

Assessing Metacognitive Knowledge Monitoring

SIGMUND TOBIAS and HOWARD T. EVERSON

College Entrance Examination Board, New York, 1996

Preparation of this report was supported by the College Board, as were Studies II-VIII. Parts of this report were prepared while the first author was a visiting faculty fellow at the Navy Personnel Research and Development Center in San Diego, California, in a program sponsored by the American Association for Engineering Education and the United States Navy.

Sigmund Tobias is Research Professor, City College of New York. Howard Everson is senior research scientist, the College Board.

.....

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

.....

Founded in 1900, the College Board is a national, nonprofit membership association of schools, colleges, and other educational organizations working together to help students succeed in the transition from school to college. The Board meets the diverse needs of schools, colleges, educators, students, and parents through the development of standards of excellence; by providing programs and services in guidance, assessment, admission, placement, financial aid, and teaching and learning; and by conducting forums, research and public policy activities. In all of its work, the Board promotes universal access to high standards of learning, equity of opportunity, and sufficient financial support so that every student has the opportunity to succeed in college and work.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is \$15.00.

Copyright © 1996 by College Entrance Examination Board. All rights reserved. College Board, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.

Printed in the United States of America.

Contents

<i>Abstract</i>	1
<i>Introduction</i>	1
<i>Knowledge Monitoring Ability and Reading Comprehension</i>	3
Study I. Estimates of Word Knowledge and Reading Comprehension.....	3
Study II. Declarative Word Reading Ability, Knowledge Monitoring and Reading Comprehension.....	5
<i>Knowledge Monitoring Ability and Classroom Learning</i>	6
Study III. Knowledge Monitoring Ability and Achievement in College.....	7
Study IV. Predicting College Achievement from KMA scores	12
Study V. Knowledge Monitoring Ability and Learning among Vocational High School Students.....	14
Study VI. Knowledge Monitoring Ability among High School Dropouts	15
<i>Prediction of Performance and Metacognitive Knowledge Monitoring Ability</i>	15
Study VII. Estimates of Performance and Predicting Scores on Standardized Tests	16
Study VIII. Knowledge Monitoring Ability and Estimates of Academic Achievement	17
<i>Metacognitive Knowledge Monitoring Ability and Mathematics</i>	20
Study IX. Mathematical Problem Solving among Elementary School Students.....	20
Study X. Relationship of Knowledge Monitoring Ability in Mathematics to Age and Achievement	21
<i>Metacognitive Knowledge Monitoring Ability and Affect</i>	22
<i>Metacognitive Knowledge Monitoring Ability and Other Variables</i>	27
Study XI. Knowledge Monitoring Ability and Need for Feedback	27
Study XII. Differences in Knowledge Monitoring Ability among Learning-Disabled and Hyperactive Students.....	28
<i>General Discussion</i>	29
<i>Endnotes</i>	34
<i>References</i>	35

Figures

1. Sample of word list for knowledge monitoring procedure.....	8
2. Sample of vocabulary items for the knowledge monitoring assessment procedure.....	9
3. Sample of expository reading passage for knowledge monitoring assessment procedure.....	10
4. Interaction of English GPA groups, hits, and test administrations	13
5. Differences between reading passage and SVT groups on knowledge monitoring ability scale	16
6. Differences between reading passage and SVT groups on raw scores for explicitly and implicitly defined words.....	17
7. Relationships between mathematical knowledge monitoring scores and grade level	21
8. Relationships between mathematical knowledge monitoring scores and math achievement group	22
9. Relationships between mathematical knowledge monitoring scores, interesting or control materials group, and math achievement group.....	26

Tables

1. Zero-Order Correlations for Selected Variables with DTLS Reading Comprehension Scores.....	4
2. Beta Weights and Associated t Test Results for All Effects on All Derived Scores	5
3. Correlations Between Knowledge Monitoring Scores, Raw Scores, and Overall Grade-Point Average in Different Subject Areas	11
4. Correlations Between Knowledge Monitoring Scores and Achievement in Mathematics	20
5. Correlations Between Knowledge Monitoring Scores and Anxiety in Mathematics	23
6. Correlations Between Knowledge Monitoring Scores and Need for Feedback	28
7. Summary Comparing Knowledge Monitoring Scores (KMA) and Raw Scores	31

Abstract

This report describes 12 studies dealing with the knowledge monitoring component of metacognition. It is assumed that knowledge monitoring is basic to other metacognitive activities, such as evaluating learning, selecting appropriate strategies, or planning, because distinguishing between what students know and do not know ought to be a prerequisite for these other higher level activities. The 12 studies, 10 in the verbal domain and two in mathematics, used various versions of a knowledge monitoring assessment (KMA) which evaluates the discrepancy between students' estimates of their knowledge in a domain and their demonstrated knowledge in that domain based on performance on a multiple choice test. The results provide a good deal of support for the construct validity of the KMA and suggest that it has considerable generalizability over different types of content and varying student populations. Since the KMA may be group or computer administered and is objectively scored, it has substantial advantages over other means of evaluating metacognition. Suggestions for further research using the procedure are made.

Introduction

Metacognition has been defined as the ability to monitor, evaluate, and make plans for one's learning (Flavell, 1979; Brown, 1980). Research has shown that learners with effective metacognitive skills are more capable of making accurate estimates of what they know and do not know, of monitoring and evaluating their ongoing learning activities, and of developing plans and selecting strategies for learning new material. A large body of literature has addressed differences in metacognitive abilities between learning disabled and regular students, as well as between generally capable learners and their less able counterparts (Schraw, in press). This research clearly indicates that metacognitive abilities are critically important for effective learning.

Metacognitive processes are usually divided into three components (Pintrich, Wolters, & Baxter, in press): knowledge about metacognition, monitoring of metacognitive processes, and control of those processes. The research described in this report concentrates on the monitoring component of metacognition, specifically, students' ability to monitor their learning by differentiating between the known and unknown. It is assumed that effective control of learning cannot occur in the absence of accurate monitoring. If students cannot distinguish between what they know and do not know, they can hardly be expected to exercise control over their

learning activities or to select appropriate strategies to attain their goals.

Our concern with assessing the ability to monitor knowledge is based on the reasoning that it is a crucial component in most learning and instructional contexts. In situations where students have to master a great deal of new knowledge, those who can accurately distinguish between what they have already learned and what is yet to be acquired have an important advantage, because they can skip over material that has already been mastered, or merely briefly review it. Such students can then devote most of their time and energies to new, unfamiliar material. In contrast, those with less adequate knowledge monitoring processes are likely to allocate their time and resources less effectively and to spend valuable time studying what is known at the expense of unfamiliar material, and consequently, to have greater difficulty mastering new subjects. For these reasons, the program of research described in this report concentrated on the development of a procedure to assess students' ability to monitor their knowledge and to differentiate between what they believe they know and do not know and what they actually know and do not know.

The purposes of this report are both to describe the metacognitive knowledge monitoring assessment (KMA) we have developed and to report on a program of research—12 studies in all—that relate scores on the KMA procedure to reading comprehension, problem solving in mathematics, and, more generally, to learning in classroom settings. In addition, we report analyses of scores on the KMA to such variables as anxiety, interest, and need for feedback, and examined the usefulness of the procedure in distinguishing learning-disabled and attention-deficit hyperactive students from those without special educational needs. All the studies described in this report used the KMA, a procedure that may be administered as either a paper-and-pencil or a computer-based assessment. Unlike other assessments of metacognitive processes, the KMA is objectively scored and does not rely on self-reports of cognitive processing.

Assessing Metacognition

Despite its importance in meaningful human learning, the assessment of metacognition has proven to be both difficult and time-consuming (Pintrich, et al, in press). Metacognition, an executive process (Borkowski, Chan, & Muthukrishna, in press), monitors and coordinates the cognitive processes employed during learning, so, as might be expected, there are considerable difficulties in assessing such higher-level processes. Metacognition is usually assessed in two principal ways: by observing students' performance or via self-report inventories. Problems associated with each of these forms of assessment are discussed below.

Observation and Verbal Reports

Assessing metacognition via observation or verbal reports usually requires all of the following:

- 1) that students work on some task individually;
- 2) that their performance is carefully observed; and
- 3) that it is recorded in some way (notes taken by observers or audio/videotapes).

Often additional steps are required before metacognition can be effectively rated, including detailed interviews with students, the development of "think aloud" protocols collected as students work on a learning task, and the recording of students' introspective reports. Multiple raters are usually required to inspect both the performance records and the interviews or introspection protocols before an effective rating of metacognition can be made (Meichenbaum, Burland, Gruson, & Cameron, 1985). Referring to this approach, Royer, Cisero, and Carlo (1993) noted that, "The process of collecting, scoring, and analyzing protocol data is extremely labor intensive" (p. 203). The resources for such efforts are rarely available in most instructional settings, and even in many university-based research programs. Pressley's work (Pressley & Afflerbach, in press, 1995) provides a good example of the complexities of conducting protocol analysis and Baker and Cerro (in press) also discuss problems with this approach, especially when it involves assessing metacognition through a process of error detection.

Labor-intensive practices such as those described above make it difficult to evaluate metacognition in many instructionally relevant settings, including secondary and postsecondary institutions, as well as training environments in business and industry, government agencies, and the military. In view of these difficulties, it is not surprising that most metacognitive research is conducted in elementary and some secondary school settings where the time of those participating in the research can be diverted for the research effort. Of course, substantial resources still have to be devoted to enable researchers to collect such data.

Self-Reports

A number of self-report measures of metacognition (Everson, Hartman, Tobias, & Gourgey, 1991; Jacobs & Paris, 1987; O'Neil, 1991; Pintrich, Smith, Garcia, & McKeachie, 1991; Schraw & Dennison, 1994) are in widespread use. Such questionnaires have the advantage of being easily administered to groups and may be scored rapidly and objectively. Self-report scales usually ask respondents to select from a set of printed choices the cognitive processes and strategies they use while learning. Such scales put a premium on effective reading abilities and, therefore, are usually not suitable for use with young children.

Unfortunately, the use of self-report measures in assessing a complex process such as metacognition raises a variety of questions, including the following: Are students aware of the cognitive processes they use during learning? Further, are students able to describe and report on the processes used, even by merely selecting from available multiple-choice alternatives? Finally, there is the question of whether students report honestly on the processes they use. While the truthfulness of students' answers is always an issue with any self-report, it may be of particular importance with respect to reports of cognitive processes used during learning, because students at any level are probably reluctant to admit that they may be relatively casual in their attempts to complete school assignments. Of course, these concerns are minimized if appraisals of any construct, metacognition in particular, do not rely on self-reports.

Rationale for Assessing Knowledge Monitoring Ability

Each of the studies reviewed below employed a technique for assessing knowledge monitoring ability that simultaneously evaluated students' self-reports of their declarative word knowledge, or their procedural problem-solving ability in math, and their demonstrated knowledge or ability. The basic strategy is to assess knowledge monitoring processes by evaluating the discrepancy between students' estimates and their actual (determined by performance on a test) knowledge or ability. The KMA was applied to the domain of students' declarative word knowledge in 10 of the 12 studies described in this report. This domain was selected because of its relevance to learning in a classroom setting. To demonstrate that the procedure generalizes to other academic domains such as math or science, two studies dealt with students' procedural knowledge in solving mathematical problems, another important domain in classroom learning at all levels.

On the KMA, students are first asked to estimate their knowledge of words or their ability to solve math problems. Actual word knowledge or problem-solving ability is subsequently assessed by administering an objectively scored test, most frequently in multiple-choice format. The discrepancies between students' estimates and their actual knowledge are used as an index of the accuracy of students' ability to monitor metacognitive knowledge.

The KMA generates four scores that reflect the relationship between students' estimates of their knowledge and their test performance. Two scores compare students' estimates of their knowledge or ability to solve a problem, and whether:

- 1) they answered the question correctly on a test (abbreviated as ++),
- 2) or answered it incorrectly (+-).

Two further scores indicate that students estimated that they did not know an item or were unable to solve a problem and

3) answer it correctly (-+) or

4) incorrectly (--).

Of course, the ++ and -- scores are assumed to reflect accurate knowledge monitoring estimates, and the +- and -+ scores inaccurate estimates.

As is true of other types of metacognitive measures, the KMA estimates also consist, in part, of self-reports. However, because of their reliance on working memory, the information in such reports typically is more readily available to students than when evoked through questions appearing on self-report inventories that often require recollections of the cognitive processes engaged in during learning, and/or how frequently the processes were used. More important, the KMA also incorporates students' actual performance on a test. Since estimated and actual performance can both be scored objectively, the procedure has a clear-cut advantage over asking students to report on their cognitive processes either in the form of protocols or by choosing alternatives on self-report inventories.

Classroom assessments are often used to determine whether students have learned the material presented in class. However, it is also important to evaluate students' ability to improve their knowledge and make accurate metacognitive estimates of whether the new material has been learned, in addition to assessing whether they have retained prior learning. Consequently, several of the studies reported below also examined students' accuracy in assessing whether they had mastered the instructional materials presented in the research.

Finally, the research described below examined the relationship of KMA scores and measures of reading comprehension, classroom learning, anxiety, interest, and need for feedback, as well as whether the KMA distinguished between students diagnosed as being either learning disabled or having an attention-deficit disorder and those not having special educational needs.

Reports of the studies are organized according to the variables examined. Because a number of the investigations dealt with multiple variables, some studies appear under more than one heading. In such instances, a detailed report of the study is given when it is first described, and the reader is directed to that description in subsequent, briefer references.

Knowledge Monitoring Ability and Reading Comprehension

A good deal of research has demonstrated that word knowledge or vocabulary is one of the major components

of reading comprehension and learning generally (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). However, few investigations have examined whether the accuracy of students' estimates of their word knowledge is an important predictor of ability to learn. If students are unable to accurately differentiate between the words they know and do not know, they will find it difficult to determine whether to slow down while reading and try to figure out the meaning of a word from the context, or go to a dictionary to define it, or go on in uncertainty or in the possibly mistaken belief that they understand the word's meaning. Such uncertainty will reduce reading comprehension among students with inadequate knowledge monitoring ability. On the other hand being able to accurately distinguish between words they can define correctly and those they cannot should enhance students' reading comprehension and their ability to learn new material. Because a great deal of research on metacognition has dealt with reading comprehension, the first two studies used the KMA's relationship to measures of reading comprehension as a criterion for assessing the validity of the KMA.

Study I. Estimates of Word Knowledge and Reading Comprehension¹

In view of the demonstrated relationships between metacognition and reading comprehension, it seemed important to evaluate the accuracy of students' monitoring of their word knowledge within the context of reading. This was expected to increase the relevance of the KMA to classroom learning. It was also anticipated that the ability to learn new vocabulary would be an important skill for reading specifically, and classroom learning generally. Furthermore, students' ability to make accurate metacognitive assessments of whether they have actually learned the meanings of new words also seems to be an important indicator of reading comprehension. Therefore, students' ability to improve their knowledge and to make metacognitive estimates of that improved knowledge were also assessed in this study.

Participants and Procedures

A total of 169 freshmen at a large urban university participated in this study. The students, considered to be at risk of doing poorly in college, attended a summer session program designed to familiarize them with the university and the skills needed to succeed in their studies. Participants were randomly assigned to one of two groups. The first group of 82 students was asked to read a 750-word passage and then complete a word list and vocabulary test composed of words that had been defined explicitly or implicitly in the text. The passage described the incidence and prevalence of heart disease, the risk factors for developing heart ailments, the technical terms for varying

degrees of the illness, the characteristics differentiating the varying degrees, and a number of ways by which the risks of developing heart disease could be reduced. Earlier research (Tobias, 1989; 1969) indicated that there was a good deal of variability in participants' prior knowledge of this material. The second was administered the Sentence Verification Test (SVT) (Royer, Lynch, Hambleton, & Bulgarelli, 1984) rather than reading the passage.

All participants were asked to indicate, by checking off one of two blanks on the word list, whether they knew or did not know each of 33 words. All of the words were defined, either explicitly or implicitly, in the passage on heart disease previously administered to the first group. When the word list was completed, students took a four-choice vocabulary test containing all 33 words on the list with instructions to select the correct synonyms or definitions of the words. To determine the participants' level of reading comprehension, the Descriptive Test of Language Skills, Reading, and Comprehension (DTLS) (College Board, 1979) was administered.

The reading passage, word list, and vocabulary test were examined by four raters to ascertain whether the words appearing in the word list were defined implicitly or explicitly in the text. The passage was revised until consensus was reached among the judges. Of the 33 words, the ratings indicated that 25 were defined implicitly (e.g., "*Epidemiologists* who have compared the prevalence of heart disease in the United States and in other countries...") and eight words were defined explicitly (e.g., "*Coronary* or heart disease...").

Results and Discussion

The accuracy of students' estimates of their metacognitive word knowledge was assessed by comparing those estimates with students' subsequent performance on the vocabulary test. The four scores described earlier were generated. Terms on the word list checked as known were scored:

- 1) correct (++) , or
- 2) wrong (+-) on the vocabulary test.

Two further scores described terms students checked as unknown on the word list and answered,

- 3) correctly (-+), or
- 4) incorrectly (--) on the vocabulary test.

These four KMA scores were computed for the total set of words and separately for those words defined explicitly and implicitly in the passage. To examine the relation between KMA scores and reading comprehension, the correlations between these scores and the DTLS subtest scores were computed and are shown in Table 1 for the entire sample, as well as separately for the group taking the Sentence Verification Test (SVT) and the group reading the heart disease passage.

The correlations in Table 1 indicate that, as expected, accurate metacognitive estimates about the number of words students thought they knew, and answered correctly on the test, (T11), had a substantial positive relationship with reading comprehension. Estimates of the number of words thought to be unknown, and answered incorrectly, (T22), were negatively related to reading

Table 1
Zero-Order Correlations for Selected Variables with DTLS Reading Comprehension Scores

	<i>Entire Sample</i>	<i>SVT Group</i>	<i>Reading Passage Group</i>
T++	.4655**	.2913*	.6474**
T--	-.4330**	-.3721**	-.5442**
T-+	-.1803	-.0885	-.2600
T+-	.0678	.2027	-.0825
E++	.3263**	.0808	.5221**
I++	.4662**	.3185*	.6302**
E--	-.3349**	-.2894*	-.4196**
I--	-.4413**	-.3822**	-.5438**
E-+	-.1390	-.1715	-.1151
I-+	-.1626	-.0523	-.2827*
E+-	.1586	.3295*	.0389
I+-	.0140	.1095	-.0877

T = Total score on word list task.

E = Words defined explicitly.

I = Words defined implicitly.

++ = Words Students claimed to know and got right on vocabulary test.

-- = Words Students claimed not to know and got wrong on vocabulary test.

+ - = Words Students claimed to know but got wrong on vocabulary test.

- + = Words Students claimed not to know but got right on vocabulary test.

* $p < .01$

** $p < .001$

comprehension. Furthermore, and also anticipated, accurate estimates of words defined explicitly (E++ and E--) and implicitly (I++ and I--) were also significantly correlated with reading comprehension, whereas incorrect estimated (E-+, E+-, I-+, and I+-) were not. The magnitude of many of the correlation coefficients is especially impressive because the participants were relatively homogeneous with respect to ability—they were considered to be at risk of doing poorly in college and, therefore, advised to participate in the orientation and pre-freshman skills program offered by the university.

The relationships between KMA scores and reading comprehension were dramatically lower for the second group of students who did not read the passage and instead took the SVT. Those reading the passage had the

opportunity to learn the meanings of previously unknown words and improve their knowledge of familiar or partially known words, whereas the students who were given the SVT did not. It was expected that students who could improve their word knowledge would make more accurate metacognitive estimates than would the others. Operationally then, it was expected that group membership (i.e., reading the heart disease passage or taking the SVT) and accuracy of metacognitive estimates would have an interactive effect on reading comprehension. This hypothesis was tested using multiple regression analysis, in which a binary vector for group, the KMA score, and their product (representing the interaction term) were entered as independent variables and the reading comprehension test score was the dependent variable. The results of that analysis are shown in Table 2.

As expected, the interaction term was significant in five of the six equations shown in Table 2. These results indicated that students who could improve their word knowledge by reading the passage made significantly more accurate metacognitive estimates than did those who did not have that opportunity. This finding is not surprising since the major skill assessed for the group reading the passage was probably the ability to infer the meaning of words, surely an important component of reading comprehension. Clearly then, the opportunity to improve word knowledge and then estimate mastery of the improved knowledge increased reading comprehension.

Table 2
Beta Weights and Associated *t* Test Results for All Effects on All Derived Scores

Score	Group		Score		Group × Score	
	Beta	<i>t</i>	Beta	<i>t</i>	Beta	<i>t</i>
T++	-.10	.44	-.62	3.33**	.88	2.67**
T--	.12	.97	.04	.19	-.54	2.32*
E++	-.42	3.00**	-.34	1.22	.85	2.60*
E--	.07	.56	.03	.14	-.41	1.70
I++	-.59	3.14**	-.07	.29	.82	2.55*
I--	.10	.84	.00	.02	-.50	2.17*

- T = Results for total word list.
- E = Results for words defined explicitly.
- I = Results for words defined implicitly.
- ++ = Words students claimed to know and got right on vocabulary test.
- +- = Words students claimed to know but got wrong on vocabulary test.
- = Words students claimed not to know and got wrong on vocabulary test.
- + = Words students claimed not to know but got right on vocabulary test.

* $p < .05$

** $p < .01$

Estimates and Actual Number Correct

On the KMA, the raw score is obtained by adding the ++ and +- scores. The KMA scores described above were a function of two factors: Knowledge as reflected in the number of items students answered correctly on the vocabulary test, and estimates of knowledge as reflected by how accurately students estimated the number they would get right. One question that arises is whether students' estimates of their knowledge contributes above and beyond their actual word knowledge as reflected by the raw vocabulary score. Of course, a great deal of research has demonstrated that students' scores on vocabulary tests are highly related to reading comprehension and classroom learning generally (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). For the KMA to be useful, it should account for more variance than the number of vocabulary items students answered correctly, irrespective of their knowledge estimates. This question was examined in the first study, and in the subsequent investigations described later in this report.

The correlation between the raw score on the vocabulary test (total number of words correct) and the DTLs was .45. As Table 1 indicates, the largest correlation among the metacognitive estimates and reading ability, $r = .65$, was between the total number of words estimated to be known and actually known (T++). The difference in the magnitude of these correlations suggests that accurate estimates of word knowledge contributed variance above and beyond the total vocabulary score. When T++ was forced into a regression equation, the total number of words correct, irrespective of prior estimates, did not contribute enough independent variance to enter the equation, indicating that the accuracy of students' estimates of their improved vocabulary were more highly related to reading comprehension than their vocabulary test scores alone. The results of this first study were encouraging with respect to the construct validity of the KMA.

Study II. Declarative Word Reading Ability, Knowledge Monitoring and Reading Comprehension²

The preceding study found strong relationships between metacognitive monitoring ability and reading comprehension in general. The purpose of the second study was to determine the KMA's relationship both to prior reading ability and to some of the components of reading comprehension, such as identifying words in context, understanding meaning, and understanding the writer's tone and assumptions. We reasoned further that knowledge monitoring may be more readily measured through the use of signal detection methods (Green & Swets, 1966; Macmillan & Creelman, 1991), which sepa-

rates feelings of knowing and other long-term memory-related phenomena into signal and noise components. Therefore, a further purpose of this study was to examine whether the signal detection paradigm could define more useful measures than the KMA scores used in Study I. Finally, relationships between KMA scores and measures of test anxiety were examined in this study, but discussion of those particular findings will occur later in the report.

Participants and Procedures

The word list and vocabulary test used in the first study were administered to students, together with the worry subscale of the Test Anxiety Inventory (Spielberger, Gonzalez, Taylor, Anton, Algaze, Ross, & Westberry, 1980) and the Descriptive Test of Language Skills, Reading, and Comprehension (DTLS) (College Entrance Examination Board, 1979), which contained three subscales: identifying words in context, understanding meaning, and understanding the writer's tone and assumptions. An index of reading ability based on an earlier administration of the DTLS was obtained from the participants' school records. The participants were 117 undergraduates at a large urban university, 65 percent of whom were women.

Results and Discussion

Knowledge monitoring ability was assessed by computing "hits" (the number of words students claimed to know and subsequently identified correctly on the vocabulary test or conversely said they did not know and failed to identify correctly on the vocabulary test) and "false alarms" (the number of words students claimed to know but did not correctly identify and those they claimed not to know yet correctly identified). Using signal detection theory, these two indices were transformed into a d' index that provides an estimate of sensitivity to metacognitive monitoring and B , an index that provides an estimate of the participants' response bias. These two indices had a reliability estimate of .78 (Cronbach, 1951).

In general, more capable readers demonstrated higher levels of metacognitive monitoring ability. The correlations of knowledge monitoring ability—as measured by the d' index—with prior reading ability and the experimental measure of reading comprehension were .35, and .39, respectively. Moreover, hierarchical multiple regression analyses permitted isolation of the effects of metacognitive monitoring ability on reading test performance, once prior reading ability and anxiety were controlled statistically. These analyses suggested that metacognitive monitoring ability was positively related to reading test performance ($B = .17, t = -2.23, p = .03$). Similarly, the correlations with the subscales on the reading test measuring vocabulary in context, literal interpretation of text, and understanding the writers' tone and assumptions were .32, .43, and .26, respectively.

Contrasts with Study I

In Study II, the reading passage in which all of the vocabulary words were defined was not administered. The correlation of .35 for the d' score was similar to the correlation of .29 (see Table 1) found in the first study between T++ and reading comprehension for those students who did not read the passage. Of course that relationship is much weaker than the correlation of .65 found in Study I between the same variables for students reading the passage. Clearly, then, these two studies suggest that the metacognitive word knowledge scores derived from the KMA had a strong, consistent relationship with standardized measures of reading comprehension and, further, that the opportunity to improve word knowledge and reestimate mastery of the improved knowledge increased the relationships with reading comprehension.

Knowledge Monitoring Ability and Classroom Learning

The first two studies were encouraging with respect to the relationship of word knowledge monitoring to reading comprehension. The results of these investigations indicated that students' metacognitive estimates of their word knowledge were closely related to competence in the domain in which the estimates were obtained, i.e., reading. One purpose of the studies to be described below was to examine whether the KMA score was related to more general academic domains such as classroom learning. The expectation of a relationship with classroom learning seemed reasonable because being able to accurately estimate one's knowledge should make it easier to acquire the large amounts of new information taught in such settings.

Four studies attempting to assess the relationship of KMA scores to classroom learning will be described below. Furthermore, because the vocabulary test and reading passage used in this set of studies dealt largely with familiar material using a minimum of technical vocabulary, the task of inferring the meanings of unknown words from the passage or estimating word knowledge seemed similar to learning in courses that rely largely on conventional vocabulary, rather than those that introduce a large set of new technical terms. Therefore, it seemed likely that declarative word KMA scores should be more closely related to students' learning in English and humanities courses than in science and social science courses.

Another purpose of the succeeding studies was to extend the research on metacognitive knowledge monitoring ability to learning in secondary and postsecondary institutions. As mentioned above, much of the research dealing with metacognition has been conducted in elementary

schools and to a lesser degree in secondary or postsecondary settings. Two of the succeeding studies examined the relationship of knowledge monitoring ability to students' overall achievement in college, and to learning in different content domains, while two others examined these parameters among high school students and those who had dropped out of school.

Study III. Knowledge Monitoring Ability and Achievement in College³

Students acquire a great deal of new knowledge during their secondary and postsecondary educational experiences. Therefore, their ability to estimate whether they have mastered either previously learned content or new material seemed to be an important characteristic of effective learners, especially in college. Accurate monitoring should enable students with effective knowledge monitoring strategies to concentrate on new materials and skim over familiar content. On the other hand, students with less effective knowledge monitoring ability may waste time practicing or reviewing what they already know, rather than zeroing in on new material or updating partially learned content. Therefore, Studies III and IV asked students to estimate their vocabulary knowledge twice: the first time to assess their prior learning and the second to determine their ability to improve prior learning. On the basis of the first study, presented earlier, we presumed that improved estimates of word knowledge would be more closely related to learning in college, as reflected in their grade point averages (GPA), than would estimates derived from prior learning.

The word list, vocabulary test, and reading passage used in the two studies reported above contained more explicitly defined words than implicitly defined words. However, it was assumed that implicit definitions might be especially important at the college level, where students frequently have to infer the meanings of new words from context. Therefore, all the materials were modified to increase the number of implicitly defined words.

Participants and Procedures

The sample consisted of 139 students attending a large urban university, though only 84 subjects completed all the materials during two sessions. Part of the sample consisted of students entering the nursing program ($N = 47$, $N = 33$ with complete data) who were taking an orientation course in nursing. The rest of the sample consisted of freshmen ($N = 92$, $N = 51$ with complete data) taking a freshman orientation course.

The word list, vocabulary test, and reading passage were revised to contain an equal number of target words that were defined explicitly and implicitly in the passage. The expository passage used in one of the previous studies was revised and a narrative version of the same passage

was developed to examine the effect of situational interest on metacognitive knowledge monitoring ability. (Findings dealing with interest will be discussed later in this report.)

The word list and vocabulary test were also revised and contained 38 words, half explicitly defined and half implicitly. Types of definitions were determined by two independent judges who rated all words. Disagreements were resolved by revising the passage until agreement was reached. Because these materials will be used in six of the studies described later in this report, a sample, consisting of the first page of the materials, is shown in Figures 1 to 3.

The word list and vocabulary test (α reliability = .80) were administered in the first session. Students were then randomly assigned to read one of two versions of the text in the second session, followed by a readministration of the word list and vocabulary test. This took place during students' classes in the presence of their instructors.

Results and Discussion

The correlation between total score on both administrations of the vocabulary test, based on the 84 students who completed both administrations of the test, was .75. This is not a test-retest reliability coefficient because students read the passage from which the meaning of the words could be inferred immediately before the second administration of the vocabulary test.

Students' estimated word knowledge and performance on the vocabulary test were determined for both administrations. Two scores were computed for each administration: the total number of correct (words in the ++ and -- categories) and incorrect (+- and -+ categories) estimates. Preliminary analysis revealed no differences between students assigned to the expository or narrative text versions, or between explicitly and implicitly defined words; therefore the data for both versions of the text and both types of words were pooled. The correlations between the correct and incorrect estimates for both administrations and students' GPAs in English, humanities, sciences, social sciences, and combined GPA were computed and are shown in Table 3. The overall GPA for the participants who were freshmen in their first term of college was based on an average of 12.1 credits ($SD = 5.6$), whereas the nursing students had a mean of 56.4 credits ($SD = 28.3$). Therefore, the correlations are presented for each group separately, as well as for the total sample. Table 3 also shows the correlations between metacognitive knowledge estimates and raw score, and the number correct on the vocabulary test.

The correlations shown in Table 3 are generally positive and frequently significant, even though they ranged in magnitude from low to moderate. The results support the concurrent validity of the KMA procedure with respect to its relationship to learning in college. As expected, correlations between knowledge monitoring scores and GPA in

Figure 1. Sample of word list for knowledge monitoring procedure.

Please indicate whether you know, or do not know each of the words listed below, by checking the appropriate space.

Abuse	___ Know	___ Do not know ___
Acute	___ Know	___ Do not know ___
Ascribed	___ Know	___ Do not know ___
Attenuate	___ Know	___ Do not know ___
Attributed	___ Know	___ Do not know ___
Benign	___ Know	___ Do not know ___
Cholesterol	___ Know	___ Do not know ___
Coronary	___ Know	___ Do not know ___
Deterrent	___ Know	___ Do not know ___
Diagnosis	___ Know	___ Do not know ___
Efficacy	___ Know	___ Do not know ___
Emanating	___ Know	___ Do not know ___
Entity	___ Know	___ Do not know ___
Epidemiology	___ Know	___ Do not know ___
Esoteric	___ Know	___ Do not know ___
Etiology	___ Know	___ Do not know ___
Fatalities	___ Know	___ Do not know ___
Genre	___ Know	___ Do not know ___
Gravity	___ Know	___ Do not know ___
Guarded	___ Know	___ Do not know ___
Implicated	___ Know	___ Do not know ___

Figure 2. Sample of vocabulary items for the knowledge monitoring assessment procedure.

For each word check the space which means most nearly the same thing as the first word.

- | | | |
|--|---|---|
| 1) Prevalent
___a) stronger
___b) winning
___c) frequent
___d) prior | 6) Ascribe
___a) refer
___b) written
___c) question
___d) bed | 11) Infarction
___a) tooth decay
___b) particle
___c) rule violation
___d) muscle death |
| 2) Attributed
___a) caused
___b) ovation
___c) stream
___d) tax | 7) Transitory
___a) move
___b) temporary
___c) carry
___d) train | 12) Fatalities
___a) fatty tissue
___b) deaths
___c) fateful
___d) take in stride |
| 3) Optimal
___a) best
___b) opening
___c) eyeball
___d) cheerful | 8) Median
___a) stripe
___b) divider
___c) middlemost
___d) negotiate | 13) Incidence
___a) new cases
___b) an example
___c) exciting
___d) event |
| 4) Obesity
___a) listen
___b) fat
___c) apology
___d) obsolete | 9) Ingest
___a) joke
___b) eat
___c) enter
___d) exit | 14) Attenuate
___a) listen
___b) reduce
___c) pay attention
___d) try |
| 5) Acute
___a) pretty
___b) serious
___c) heavy
___d) often | 10) Residual
___a) lasting
___b) live
___c) income
___d) clever | 15) Guarded
___a) uncertain
___b) optimistic
___c) degrees
___d) watchful |

Please turn to next page to continue

Figure 3. Sample of expository reading passage for knowledge monitoring assessment procedure.

Read this passage carefully:

Coronary or heart disease is a major health problem among all ethnic, racial and occupational groups in the United States. In addition to coronary disease, health workers are worried about many other maladies affecting Americans, such as cancer, AIDS, and other equally serious conditions. However, compared to all other serious illnesses, coronary problems cause more than half of the total number of fatalities or deaths in the United States. To be exact, 55% of the deaths among all groups in this country or more fatalities than for all the other illnesses combined, may be ascribed to coronary disease. Not only is coronary disease responsible for the greatest number of fatalities in this country, but it is also the most prevalent, or frequent, of all the serious illnesses. That is, coronary disease is more prevalent than all the other serious conditions combined.

The incidence, that is the number of new cases, of coronary disorders, is higher among men than among women for the country as a whole. The incidence of heart disorders is also higher for cigarette smokers than it is among non-smokers. A higher incidence of coronary disease among Americans is also attributed to alcoholism, drug addiction, and tobacco. The etiology, or causes, of coronary disease among Americans are not completely clear, but excessive use or abuse of alcohol and the other substances mentioned above is often linked to coronary disease. In addition, tension, air pollution, weighing too much, and engaging in too little exercise are also implicated as causes of heart disease among people living in the United States.

The gravity of heart disease for people in general is a function of the magnitude of coronary damage. The heart is basically a muscle similar to all the others in the human body. The amount of damage to the heart muscle, or myocardium, determines the seriousness of the illness. The most serious type of damage, which is called myocardial infarction, occurs when the heart muscle dies. One major difference between the myocardium and other muscles in the human body

Table 3

Correlations Between Knowledge Monitoring Scores, Raw Scores, and Overall Grade-Point Average in Different Subject Areas

<i>Variables</i>	<i>Administration 1</i>			<i>Administration 2</i>		
	<i>Group</i>	<i>Correct Estimate</i>	<i>Raw Score</i>	<i>Correct Estimate</i>	<i>Raw Score</i>	<i>Raw Score</i>
Total GPA	<i>N</i>	<i>r</i>	<i>r</i>	<i>N</i>	<i>r</i>	<i>r</i>
Total	101	.20*	.01	94	.09	-.00
Freshmen	65	.09	-.25	61	-.10	-.21
Nurses	36	.28*	-.37*	33	.19	.17
<i>English GPA</i>						
Total	72	.30**	.19	63	.19	.05
Freshmen	53	.31**	.10	48	.00	.16
Nurses	19	.25	.33	19	.45*	.44*
<i>Humanities GPA</i>						
Total	82	.26**	.04	74	.13	.00
Freshmen	52	.12	-.21	46	-.11	.22
Nurses	30	.48**	.40*	28	.35*	.24
<i>Science GPA</i>						
Total	65	.18	-.01	63	.03	-.07
Freshmen	28	.11	-.30	27	-.28	-.47
Nurses	37	.26	-.42*	36	.18	.26
<i>Social Science GPA</i>						
Total	64	.18	.26	63	.24*	-.26*
Freshmen	26	.15	.10	29	.14	.18
Nurses	38	.09	-.31.00	34	.14	.10

* $p < .05$ ** $p < .01$

English were generally highest; presumably the ability to accurately estimate word knowledge is more important in English than in other subjects. Relationships with humanities GPAs and with the combined GPA were generally significant but lower than those with English GPAs; correlations with social science and science GPAs were generally lower and usually not significant. The largely nonsignificant relationships with social and behavioral science GPAs were surprising, because it had been assumed, perhaps naively, that these courses would present fewer technical terms and unfamiliar vocabulary than the natural science courses. Perhaps grades in these courses, as with those in science, reflect greater domain-specific knowledge than is true in English and humanities courses.

The significance of the correlations reported in Table 3 varies widely, probably as a function of at least three factors. First, there is a different number of cases in each cell because some students were not present for both administrations of the materials, leading to variability in the predictors. Second, college grades are often unreliable (Werts, Linn, & Joreskog, 1978; Willingham, Lewis,

Morgan, & Ramist, 1990), reducing the magnitude of any correlations with them. Third, students completed varying numbers of courses in each area, thus GPAs may have been based on one or a few courses in some fields, reducing the stability of the criterion. The reliability of the grades may have been reduced further by three factors:

- 1) students took dissimilar courses in each of the broad subject areas shown in Table 3;
- 2) when similar courses were taken, they were taught by different instructors; and
- 3) there were differences in students' major fields of study.

As expected, the correlations between knowledge monitoring scores and grades in English were generally higher and more frequently significant than was true of any other subject. For the 84 students for whom there were complete data for both administrations of the vocabulary test, the mean total score increased from 23.3 (SD = 6.0) for the first vocabulary test to 26.0 (SD = 6.6) for the second [$t(83) = 5.53, p < .001$]. Thus students clearly learned the meanings of some words after having the chance to improve their word knowledge by reading the passage.

However, in contrast to the results of the first study, the relationships between the metacognitive scores and grades shown in Table 3 were generally higher before students read the passage rather than afterward. The Study I findings of higher relationships with DTLs scores on the second administration of the KMA may be attributable to the use of reading comprehension scores rather than grades as criteria. Apparently inferring the meaning of words is a more important component of reading comprehension than of classroom learning more generally.

As mentioned, although it was assumed that having the chance to improve word knowledge before estimating it would be more similar to the way in which students learn in their courses than would the process of merely estimating prior word knowledge, the relationships with grades were not higher for the second administration of the KMA than for the first. While the increase in vocabulary score after the passage was read was statistically significant, fewer than three new words were learned from the passage. Perhaps such modest acquisition of vocabulary does not reflect the amount of learning that takes place in college courses, leading to lower relationships with metacognitive monitoring scores on the second administration. The degree of similarity between the knowledge monitoring task and classroom learning might have been greater if students had been instructed to study the passage more intensely or asked to pay special attention while reading words they had previously seen on the KMA word list. Such instructions might have increased the correlations with GPA for the second administration. It remains for further research to explore that possibility.

Table 3 also indicates that the correlations with the number correct on the vocabulary test were generally similar to the relationships with correct estimates of word knowledge. Due to the varying Ns in the different cells, the significance of differences in correlations was examined using a *t* test developed by Hotelling (1940). Relationships with GPA based on the total group indicated that seven correlations with knowledge monitoring scores were higher than similar correlations with raw scores (one difference was significant at $p < .05$) while the correlations based on raw scores were higher three times (none significantly so). For freshmen, the correlations with knowledge monitoring scores were higher twice, though not significantly so, than those with raw scores while the correlations with raw scores were higher eight times (two significant at $p < .05$). Finally, for nursing students, correlations based on knowledge monitoring scores were higher five times (none significant), while relationships based on raw scores were higher five times (one significant at $p < .05$). Thus, the knowledge monitoring scores appeared to add little independent explanatory power to the relationship with grades beyond that accounted for by the number correct on the vocabulary test.

The findings of this study, in contrast to the findings from the first two investigations, suggest that estimates of knowledge seem to account for little independent variance in GPA above that attributable to the number correct on the vocabulary test. Conceivably the low reliability of college grades (Werts, Linn, & Joreskog, 1978; Willingham, Lewis, Morgan, & Ramist, 1990), referred to above, may have contributed to these findings. The criterion in the first two studies consisted of test scores, which are much more reliable than grades.

Study IV. Predicting College Achievement from KMA Scores

The preceding study dealt with concurrent validity in relating knowledge monitoring scores to students' achievement in college. The fourth study investigated the predictive validity of KMA scores by examining whether metacognitive estimates of knowledge would predict entering students' performance during their first year of college.

Participants and Procedures

The materials used were identical to those described in Study III. They were administered while students attended a pre-freshman skills program before beginning their first semester of college. Achievement was determined by obtaining students' GPAs at the end of their first year of college in the same subjects examined in the prior study: English, humanities, sciences, and social sciences, as well as the combined GPA. The sample consisted of 115 students (59 female) participating in a skills program intended for students considered at risk of doing poorly in their first year of college.

Each participant in Study IV completed all of the study materials and took similar types of courses. High- and low-achievement groups were created by dividing students at the GPA median for the different academic areas and for the combined GPA. Then differences in knowledge monitoring ability between the groups were examined. Mixed between- and within-subjects analyses of variance were computed to determine the significance of differences between the first and second administrations, and of differences in estimates of knowledge between groups above and below the GPA median. At the conclusion of freshman year, it was determined that 95 of the 115 original participants had completed some courses at the college.

Results and Discussion

The number of correct estimates students made of their word knowledge was determined. As in the prior studies, correct estimates were defined by combining the ++ and -- categories. Preliminary analysis again indicated that there were no differences between the results obtained for

the expository and narrative passages, nor between those for the words defined explicitly or implicitly. Therefore, these data were pooled for the succeeding analyses.

ANOVA indicated that, as expected, students above the median GPA ($N = 48$) made significantly more accurate overall estimates of their knowledge [mean = 49.2, $F(1,93) = 6.42, p < .05$] on both administrations than did those below the median ($N = 47$, mean = 45.8); the size of that effect, determined by η^2 (SPSS, 1993), was .065. Also as expected, there was a significant difference between the first (mean = 22.9) and second administration (mean = 24.5) of the word list and vocabulary test [$F(1,93) = 14.95, p < .01, \eta^2 = .138$], though there was no interaction between these variables. A similar analysis was made using the number correct on both administrations of the vocabulary test as the dependent variable. The analysis indicated that the differences between the high- (mean = 43.2) and low-GPA group (mean = 39.3) on the vocabulary test were not significant [$F(1,93) = 2.73, \eta^2 = .029$], while the differences between the first (mean = 17.7) and second administrations (mean = 24.5) were highly significant ($F(1,93) = 198.04, p < .001, \eta^2 = .68$); again there was no interaction between variables.

High- and low-achieving groups in English, humanities, science, and social science courses were also identified by dividing the students at the median GPA for each of these subject areas and examining the significance of differences in the number of correct estimates of knowledge. In English, the overall differences in the accuracy of the estimates between students above (mean = 48.9) and below the median (mean = 45.4) were significant [$F(1,82) = 6.18, p < .02, \eta^2 = .07$], as were the differences between the first (mean = 45.6) and second administrations [mean = 48.7, $F(1,82) = 11.92, p < .01; \eta^2 = .127$]. Furthermore, there was an interaction between groups and administrations [$F(1,82) = 4.41, p < .05; \eta^2 = .051$]. The interaction, as shown in Figure 4, suggests that while the accuracy of both groups' estimates of known and

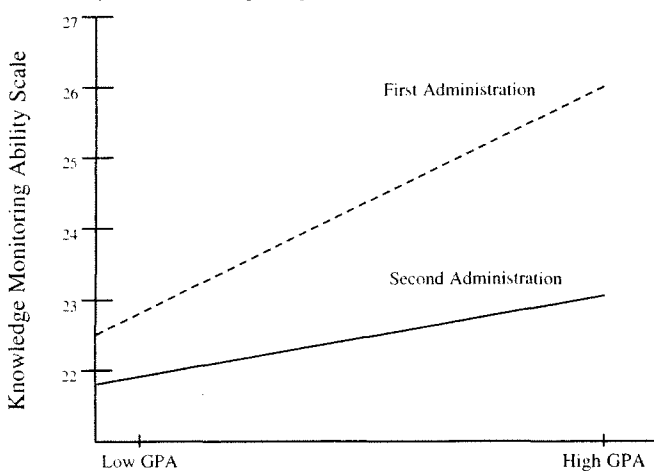


Figure 4. Interaction of English GPA groups, hits, and test administrations.

unknown words increased from the first to the second administration, higher achieving students made greater gains. A similar analysis was computed for the number correct on both vocabulary test administrations. The findings indicated that the difference between the high- (mean = 42.9) and low-GPA groups [mean = 38.9, $F(1,82) = 5.43; \eta^2 = .062$] was slightly smaller than that determined when the metacognitive knowledge scores were used, but there was a stronger effect for differences between first (mean = 18.0) and second administrations [mean = 23.6, $F(1,82) = 169, p < .001; \eta^2 = .673$]; there was no evidence of interaction in these results.

Similar analyses were made for students above and below the median GPA in humanities courses (Art, History, Music, Philosophy, World Civilization, World Humanities, and World Arts). Differences between high (mean = 49.4) and low humanities GPA groups (mean = 45.3) were also significant [$F(1,81) = 7.96, p < .01; \eta^2 = .089$], as were differences between the first (mean = 23.0) and second administrations [mean = 24.5, $F(1,81) = 9.94, p < .001; \eta^2 = .109$]; there was no interaction. The same type of analysis was also computed for the number correct on the first and second vocabulary tests; again it revealed somewhat smaller differences between the high- (mean = 43.1) and low-GPA groups [mean = 39.0, $F(1,81) = 4.18, p < .05; \eta^2 = .049$] and larger differences between the first (mean = 17.8) and second administrations [mean = 23.4, $F(1,81) = 179.2, p < .001; \eta^2 = .689$] than the results for knowledge monitoring scores. There were no significant differences between the science or social science GPA groups using either the knowledge monitoring scores or the raw scores.

The relationships between metacognitive scores and GPA were generally similar to those reported in Study III, supporting the predictive validity of the KMA scores. In contrast with the prior study, in which both KMA and raw scores had fairly similar patterns of relationship, the metacognitive scores had a significant effect on overall GPA, whereas the raw scores did not. Furthermore, the KMA scores accounted for more variance between groups than did the number correct on the vocabulary test in two of three other comparisons, supporting the construct validity of the KMA procedure.

Several factors are likely to have reduced the magnitude of the effects and the generalizability of the results to other groups of college students. As in the first study, participants in the pre-freshman program were considered to be at risk of poor performance in college. This may have reduced the range of achievement for the sample and, therefore, may also have reduced the differences in knowledge monitoring ability between the groups. Furthermore, even though data were *not* collected in sections of the pre-freshman program devoted exclusively to English as a Second Language (ESL), some of the students were enrolled in both ESL and other sections, and thus ended up

as part of the sample. The presence of non-native English speakers may also have reduced the variability among participants and narrowed group differences. Further research limited to native English speakers, who are more heterogeneous in terms of academic skills than was the present sample, is needed to determine whether knowledge monitoring ability differences between low- and high-achieving students are greater than those reported here.

In general, KMA scores seemed to more successfully differentiate the capable students, whose grades were above the median, from those less able than did the raw scores, replicating the findings of Studies I and III. The knowledge monitoring scores accounted for anywhere from 1 to 4 percent more variance than did similar analyses using the raw scores. It was also interesting that the analysis of differences in raw scores between the first and second vocabulary test administrations always accounted for substantially more variance than did a similar analysis based on knowledge monitoring scores. The latter finding is reasonable and supports the construct validity of the KMA procedure in that most students learned some new words from reading the passage, though their knowledge monitoring ability was not equally enhanced. However, it should be noted that the results for English grades indicated that there were greater increases in knowledge monitoring ability for capable students than for their less able peers (see Figure 4). These findings suggest that while all students increased both their demonstrated knowledge and their knowledge monitoring ability from first to second test administration, the increases in monitoring ability were greater for more capable students (i.e., those whose English grades were above the median). Apparently there was a greater degree of improvement in such students' metacognitive skills than in those of their less able colleagues.

It should be noted that many of the students in this sample took less than a full-time schedule of courses. This fact is likely to have decreased the reliability of the GPA, because it was based on fewer courses and credits than is usually the case after a year of college. This may also limit the generalizability of the results to other groups of students. Therefore, to increase both the reliability and variability of this criterion, it would be useful to investigate the predictive validity of the KMA procedure for a large number of full-time students.

Study V. Knowledge Monitoring Ability and Learning among Vocational High School Students⁴

College students were used as subjects in all of the previous studies. Individuals attending college are likely to be more studious and academically oriented than are students at secondary levels and therefore more likely to be reflective about what they know and do not know. Thus, one purpose

of Study V was to examine the applicability of the KMA procedure to students attending a vocational high school.

Participants and Procedures

All of the participants attended a vocational high school in a large urban school system. A total of 61 students (59 male) participated in this study. The students' ages ranged from 16–19. This study employed the word list and vocabulary test described in the two preceding studies; the reading passage was not administered. Students were tested during one of their regular school classes. In addition, test anxiety scales were administered and students were asked to estimate their grades on tests given in one of their vocational classes. Students' overall GPAs were obtained from the school's permanent records.

Results and Discussion

Students were divided into two groups at the GPA median. Two multivariate analyses of variance (MANOVA) were computed: the first examined differences between the high- and low-GPA groups in terms of the accuracy of students' estimates of their knowledge (using the ++, +-, and -- scores) and the second analysis examined group differences in students' word knowledge (the sum of ++ and -+ scores equalled the number correct on the vocabulary test). The MANOVA indicated that the overall differences in knowledge monitoring ability between the high- and low-GPA groups were significant [Wilks $F(3,57) = 3.17, p < .05$, effect size = .143]. Univariate analyses showed that only the difference between the high- (mean = 17.8) and low- (mean = 14.4) GPA groups on the ++ scores was significant [$F(1,59) = 9.35, p < .01$].

The MANOVA computed on group differences in the number correct on the vocabulary test also indicated a significant difference between the groups [Wilks $F(2,58) = 5.35, p < .01$, effect size = .156]. Univariate analyses showed that the differences in ++ scores were the same as in the preceding analysis; however, in this analysis, group differences between the high- (mean = 3.5) and low- (mean = 5.2) GPA groups on the -+ scores were also significant [$F(2,58) = 5.59, p < .05$]. As expected, the results indicated that the capable students estimated that they knew and actually knew more words than did those with lower GPAs, while the latter group estimated that they did not know more words than did the students who were above the GPA median.

The significant differences between the two GPA groups replicated the results of the two prior studies and confirm the relationships between metacognitive knowledge estimates and classroom learning. The results of the second analysis do not support the additional importance of obtaining students' estimates of their knowledge, because the differences between the GPA groups in their actual vocabulary knowledge were also significant, and

slightly greater than the differences in knowledge estimates. However, the word list and vocabulary test were developed for a college population; perhaps these materials were so unfamiliar to these high school students that their estimates were based on little more than chance.

Study VI. Knowledge Monitoring Ability among High School Dropouts⁵

The high percentage of students who drop out of school is a major problem, especially in times when even most entry-level positions in business and industry call for greater levels of skill than ever before. At a time when the advent of the information superhighway is beginning to re-define the job functions of lower- and mid-level workers, it is vital that students complete a secondary school education in order to have better chances of finding employment. Nevertheless, there are indications that the number of school dropouts is increasing. "A national estimate suggests that 25 percent of fifth graders will not make it through high school graduation" (Mann, 1986, p. 309).

There are many reasons for students' dropping out of school, but Tanner (1990) suggests that "School-based reasons are the most important self-reported explanation of dropping out for all groups of adolescents" (p. 80). Chief among these is poor performance in school. When asked why they had dropped out of school, more than one-third of the students said, "Because I had bad grades" or "because I did not like school" (Mann, 1986, p. 309). These findings are substantially similar to those reported by Ekstrom, Goertz, Pollack, & Rock (1986). It therefore seemed reasonable to assume that students who drop out of high school would have lower metacognitive knowledge monitoring abilities than other students. This assumption was examined in Study VI.

Participants and Procedures

The word list and vocabulary test employed in Studies II to V were administered, together with some test anxiety scales that will be described later in this report. The reading passage was not used.

A total of 89 subjects participated. The dropout group consisted of 42 individuals (14 female) who were attending a General Equivalency Diploma program. A group of currently enrolled students (47 total, 16 female) who had a school GPA of at least B- also participated. None of the students in the latter group had given any indication that they were at risk of dropping out of school.

Results and Discussion

Two MANOVAs, identical to those computed in the preceding study, were used to determine the significance of differences between the high school dropouts and continuing students. The first analysis found significant overall group differences [Wilks $F(3,79) = 4.08, p > .01,$

effect size = .134] in KMA scores (++ , +- , and -- scores). Univariate analyses indicated that the dropout group (mean = 12.7) differed from the continuing students (mean = 16.2) on the ++ scores [$F(1,81) = 8.83, p < .01$] and on the +- scores (mean = dropout 10.6, continuing students 8.5); [$F(1,81) = 6.11, p < .02$]. A similar analysis of actual knowledge (++ and +- scores) also indicated significant, though somewhat smaller, group differences [Wilks $F(2,80) = 4.61, p < .01,$ effect size = .103]. Univariate analyses indicated that only the difference in the ++ scores was significant; of course, the statistics for this effect were identical to those used in the first MANOVA.

The results indicate that, as expected, students who dropped out of high school had less adequate knowledge monitoring abilities than continuing students. Analysis of differences in raw scores revealed similar, though somewhat smaller, effects. The results suggest that the limited knowledge monitoring abilities of students who dropped out of school may have made schoolwork more difficult for them and contributed to poor performance, which is consistent with the descriptions of school dropouts in the literature.

Summary: Knowledge Monitoring Ability and Classroom Learning

As expected, the four studies discussed in this section found significant relationships between metacognitive knowledge monitoring scores and classroom learning. The studies used different student samples: enrolled college students, those about to enter college and enrolled in a pre-freshman skills program, vocational and regular high school students, and those who had dropped out of school. Because relationships with knowledge monitoring ability were in the expected direction for the different samples, it may be inferred that the KMA is generalizable across a variety of student groups. In most of the studies, the KMA scores accounted for more variance than did the raw vocabulary scores, supporting the construct validity of the KMA procedure.

Prediction of Performance and Metacognitive Knowledge Monitoring Ability

It was reasoned that students who were capable of accurately estimating their knowledge of vocabulary on the KMA should also be more accurate in predicting their performance on examinations related to their current studies. This section will describe three investigations examining this assumption.

There has been some research on students' predictions

of their performance in courses and on tests, though none relating the predictions to metacognition or knowledge monitoring ability. Keefer (1971) found that college students who accurately estimated their performance achieved at a significantly higher level than did those who estimated less accurately and had a more positive self-concept than their low-estimating counterparts. Holen and Newhouse (1976) found that students' predictions of their grades on a course examination correlated as highly with actual performance as did their GPAs and were significantly more accurate predictors of performance than other variables, such as grades in prerequisite courses or GPAs. Furthermore, students' performance predictions contributed significant unique variance to actual final grade, above that contributed by high school and college GPA, or grades in prerequisite courses in predicting that grade. Harris (1990) found that students who were accurate estimators of their test performance in psychology earned a significantly higher final grade in Introductory Psychology than did less-accurate estimators.

The research on prediction of performance suggests that more capable students make more accurate predictions of their performance than do their less able counterparts. Because the studies described in the preceding section found that higher KMA scores were associated with higher GPAs, the findings dealing with predictions of performance suggest that students who make accurate metacognitive assessments of their knowledge should also make more accurate predictions of their test scores.

Study VII. Estimates of Performance and Predicting Scores on Standardized Tests

In addition to describing the relationship between students' estimation of and actual performance on tests, this study also varied the reading passage used, to examine its contribution to students' estimates of their test performance. Furthermore, it was decided to examine performance on a standardized test of known reliability to reduce possible error. Studies I and II used a standardized measure of reading comprehension (the DTLs, College Board, 1979) as the criterion and the results relating test performance to KMA were more positive than were the results where less-reliable student grades were used. Therefore, a test with known reliability (.88) was used in this study.

It was expected that Introductory Psychology students who could accurately monitor their knowledge would also be more accurate in predicting their actual and estimated scores on the Advanced Placement (AP) Test in Psychology (College Board, 1988) before and after completing it, and that they would also earn higher scores on the AP test than would their peers who estimated less accurately. Finally, as suggested by other studies of students' ability to estimate their performance, it was predicted that

students with high KMA scores would expect to obtain higher grades than those with lower scores.

Participants and Procedures

A total of 77 students (41 female) taking an Introductory Psychology course at one of the campuses of a large urban university volunteered to participate in the study. Participation in this study satisfied a course requirement.

The AP Examination in Psychology (College Board, 1988) was administered to students enrolled in an Introductory Psychology course. Students received a description of the different areas covered by the AP test and were asked to predict how many of the 100 items they would be able to answer correctly before they took the test and again after the test was completed. Half of the sample ($N = 39$) was randomly assigned to read the expository version of the test passage used in two of the studies described in the earlier section, while the other half ($N = 38$) performed an unrelated task, reading the text selection titled "Teaching the Mentally Retarded" from the Sentence Verification Technique (SVT) (Royer, Carlo, Dufresne, & Mestre, 1994) and answering questions on that passage. The same word list and vocabulary test used in Studies II to VI were then administered to all participants.

Students were also asked to predict their final grade in the Introductory Psychology course they were taking. On this campus, the accuracy of their grade predictions could not be determined because regulations protecting students' privacy made it impossible to obtain that information.

Results and Discussion

More accurate KMA scores were expected for the group responding to the word list and vocabulary test after reading the passage compared to the other group who received the SVT, which was irrelevant to the task.

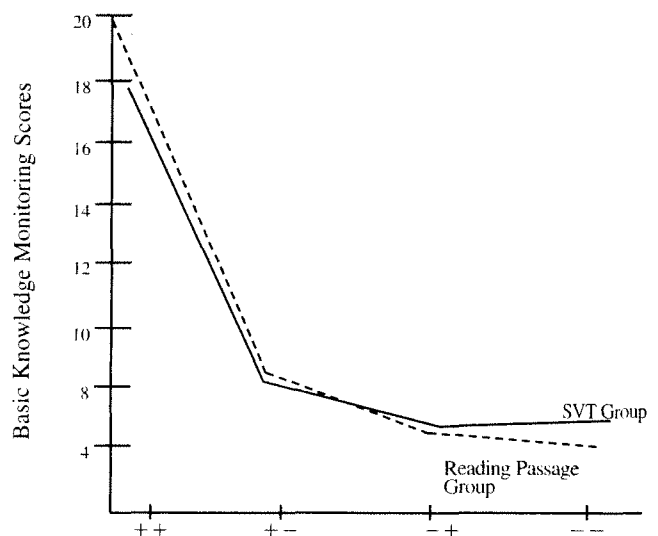


Figure 5. Differences between reading passage and SVT groups on knowledge monitoring ability scale.

Surprisingly, MANOVA based on the total number of accurate estimates ($++$ and $--$) revealed no significant differences between the groups. (See Figure 5.) Examination of the four basic KMA scores for both explicitly and implicitly defined words indicated that there were some group differences (see Figure 6), but that these differences were canceled out when the data were combined into total number of correct estimates.

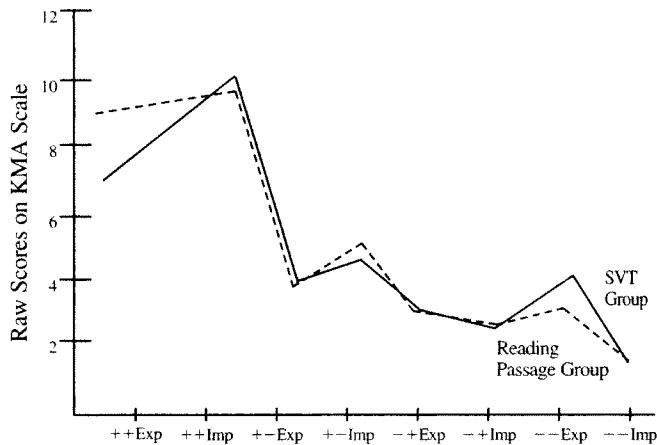


Figure 6. Differences between reading passage and SVT groups on raw scores for explicitly and implicitly defined words.

When a MANOVA was performed on six of the basic scores (the scores for the $+-$ category for explicitly and implicitly defined words were eliminated to reduce collinearity), the overall differences between the groups were significant [$F(6,70) = 3.71, p < .01$]. Univariate F tests indicated that the students who read the passage made more accurate metacognitive estimates on explicitly defined words in the $++$ category [$F(1,75) = 5.97, p < .02$] and had fewer explicitly defined words in the $--$ category [$F(1,75) = 4.74, p < .05$].

Student groups were divided at the median on total number of accurate metacognitive estimates (combining the $++$ and $--$ categories). A MANOVA was performed to examine the significance of the differences on students' predictions of their AP scores before and after they took the test, their actual AP scores, and their expected final grades in the Introductory Psychology class. There were no differences on the AP test data or on the expected final grades between the groups who read either the passage or the SVT [$F(4,69) < 1$], but differences between the high and low groups were significant [$F(4,69) = 2.83, p < .05$, effect size = .141]. Univariate tests indicated that the high knowledge monitoring group obtained higher AP scores [mean = 43.6, $F(1,72) = 7.81, p < .01$], and that differences in expected final grade in the course were just short of significance [$F(1,72) = 3.40, p < .10$]. There was no interaction between the groups who read either the passage or the SVT and the knowledge monitoring groups.

The data were also analyzed with respect to the number correct on the vocabulary test, dividing the groups at the

median and computing the significance of differences in relation to AP test scores and final course grades. The results were similar in that there were no differences between the groups who had read the passage or the SVT, but there was a significant difference between the groups above and below the median on pre-passage KMA score [$F(4,69) = 6.47, p < .01$, effect size = .27]. Univariate analysis again indicated only one significant difference on actual AP scores between the groups above (mean = 45.4) and below (mean = 34.2) the median on the vocabulary test score. Again there was no interaction among the variables. Unlike the prior studies, where differences in metacognitive knowledge estimates were usually greater than those on the vocabulary test raw score, the effect size for these data was larger using the vocabulary test results than the knowledge monitoring data (.27 compared to .14).

The results indicate that students with a high score on the vocabulary test and with high ability to monitor their word knowledge also obtained higher scores on the AP exam and, at a marginally significant level, expected higher final grades in the course. The absence of group differences on AP score predictions before taking the test was not surprising because students were unfamiliar with the test. Beyond being informed about the categories of knowledge covered, they had no information about the difficulty of the items, the types of preparation possible for the test, or specifically what they would be questioned on. The absence of differences on students' predictions after taking the test was a little more surprising, because participants now had a much clearer idea about what the test covered. Perhaps this brief exposure to the test was inadequate to familiarize them with the domain covered by the AP examination.

Study VIII. Knowledge Monitoring Ability and Estimates of Academic Achievement

Ideally, of course, students' predictions of their performance in courses for which they were registered should have been studied. As is not the case with the AP test, students should have enough information to make fairly accurate predictions about their final grades in courses based on their experience in the class and with the subject matter, the instructor, and the procedures of the course. This study was intended to examine this assumption, in addition to attempting to replicate the findings from the AP study.

Participants and Procedures

The procedures were identical to those in the previous study with two exceptions: first, the predictions students made about their final grades were compared to the actual final grades obtained in the course; second, students took 12 quizzes in this class. (The instructor used the 10 highest quiz scores to determine the final grade.) The grades on these

quizzes were available as additional dependent variables.

A total of 75 college students enrolled in Introductory Psychology participated in this study. The students received extra credit for taking part in the research.

Results and Discussion

The first set of analyses was computed to examine the consistency between the findings in this study and the preceding one. As in the prior study, a test for the significance of differences between the group who read the passage and the SVT on the ++, +-, and -- scores revealed no differences between the groups. When the component scores based on explicitly and implicitly defined words were examined, overall differences between the groups were significant [$F(6,68) = 2.57, p < .05$]. Univariate analysis indicated that the group reading the passage had fewer -- scores for the explicitly defined words [$F(1,73) = 7.69, p < .01$] and more ++ scores for the implicitly defined words [$F(1,73) = 7.29, p < .01$]. These results are identical to those in the preceding study and suggest that combining the data may have obscured existing group differences. Both sets of results point to the importance of conducting a study specifically designed to determine which set of data is the best indicator of the latent knowledge monitoring ability variable.

The analysis of differences between groups scoring high and low on knowledge monitoring ability in predicted before, after, and actual AP test scores, and final grades was also similar to that in the preceding study with one addition: students' actual final grades in the course were available as an additional dependent variable. Two groups were created by dividing students at the median on the total number of accurate estimates of vocabulary knowledge and computing a MANOVA to examine the significance of differences on the AP and grade data; nine students were eliminated due to missing information. No differences between the groups who read either the passage or the SVT were found [$F(5,58) = 1.37$]. In contrast with the prior study, the differences between the two groups in knowledge monitoring ability only approached significance [Wilks $F(5,58) = 2.21, p = .066$; effect size = .16]. Univariate analysis indicated that the high knowledge monitoring group had significantly higher AP scores (mean = 45.2) than did the lower group [mean = 36.7; $F(1,62) = 10.02, p < .01$]; there were no differences on predicted score either before or after the AP exam was taken, or on predicted and actual final grades.

The finding that the high and low knowledge monitoring ability groups differed only on actual AP test performance, rather than on any of the predictions, also replicated that of the prior study. The failure to find differences on the final-grades variable may have been a function of the limited range of the grades; A–D grades (there were no F grades in this sample) were converted to

their numerical equivalents, yielding only four scores. Furthermore, 76 percent of the grades were B or higher, further limiting their variability. The interaction between knowledge monitoring ability groups and those who read either the passage or the SVT was of borderline significance [$F(5,58) = 2.18, p = .069$], principally attributable to the fact that the low-ability group's estimates of their AP scores and their final course grades were actually higher than those of the high-ability group, while the actual scores and grades of the former group were lower than those of the latter.

An identical MANOVA was performed with students divided at the median on the number of words correct on the vocabulary test as the independent variable. There were highly significant differences between the groups [Wilks $F(5,58) = 5.70, p < .001$; effect size = .33]. Univariate analysis indicated that the group scoring high on the vocabulary test also had higher AP scores [mean = 47.0, $F(1,62) = 22.89, p < .001$] than the group scoring low (mean = 35.1). Contrary to the analysis based on predictions of knowledge, with this analysis, the high-scoring group also received higher final grades [mean = 90.4, $F(1,62) = 5.24, p < .05$] than did the low-scoring group (mean = 85). Again, the interaction between the groups who read either the passage or the SVT and the vocabulary test results approached significance [Wilks $F(5,58) = 2.12, p = .076$], attributable to the low-scoring group predicting higher AP scores both before and after, and final grades, while actually obtaining lower scores on all three measures.

The second set of analyses examined the relationship between knowledge monitoring scores and indices of in-class student performance, such as scores on quizzes and on the essay and multiple-choice parts of the final examination. Because the instructor informed students that only the 10 highest scores on the 12 quizzes given would count for the final grade, many students missed some quizzes. Therefore, for students taking at least 10 of the quizzes, the mean score on all the quizzes taken was used as one of the dependent variables. Student groups were then divided at the median on the knowledge monitoring scores and a MANOVA was performed on the quiz and final examination data; missing data limited this analysis to 70 students.

There were no significant differences on the class performance indices between the groups taking the SVT or reading the passage [$F(4,63) = 1.04$]. There was an overall significant difference between the high- and low-knowledge monitoring ability groups [Wilks $F(3,64) = 4.36, p < .01$, effect size = .17]. Univariate analyses indicated that the high-ability group had significantly higher scores on the multiple-choice part of the final examination (mean = 25.1) as compared to the low-ability group [mean = 21.2, $F(1,66) = 12.66, p < .01$]. Differences between the groups on mean quiz scores were of borderline significance [$F(1,66) = 3.02$,

$p = .09$], with the high-ability group getting better scores (mean = 4.51) than the low-ability group (mean = 4.1). (Each quiz had a total of six raw score points.) There was no interaction between knowledge monitoring ability and whether groups read the passage or the SVT.

An identical analysis was performed with students divided at the median on the number of words correct on the vocabulary test as the independent variable. There were overall differences for the high- and low-scoring groups [Wilks $F(3,64) = 6.44, p < .01$, effect size = .232]. The high-scoring group had significantly higher scores on both the essay [mean = 17.2, $F(1,66) = 7.44, p < .01$] and multiple-choice [mean = 25.5, $F(1,66) = 18.72, p < 0.01$] parts of the final exam and on the mean for the quizzes [mean = 4.6, $F(1,66) = 7.13, p = .01$] than did the lower-scoring group (means = 14.5, 20.9, and 4.0, respectively). In this study as in the preceding one, vocabulary score was more effective than knowledge monitoring ability in terms of differentiating between students on AP and final grades (.33 effect size vs. .16) and on classroom tests (.232 compared to .17 effect size).

Knowledge Monitoring Ability and Estimates of Performance among Vocational High School Students

In Study V, examining the relationship between knowledge monitoring ability and classroom learning among vocational high school students, the participants were also asked to predict their grades on a final course examination both prior to and after taking it; the actual score on that test was available as a dependent measure. MANOVAs indicated that neither the metacognitive knowledge monitoring estimates nor the raw scores on the vocabulary test were significantly related to either the predicted or actual grades. The failure to find any differences is at variance with the findings of the two preceding studies involving college students.

There are a number of differences between the studies involving vocational high school and college students, in addition to the population differences, that may account for the diverse findings. The vocational high school students were asked to predict their performance on a final exam in the class they were taking and presumably had a much better idea of the content of the exam and how to prepare for it than did the college psychology students who had very little basis for knowing what to expect on the AP test and could not prepare for it at all. Furthermore, given that the vocational students had been graded on other exams in that class, they—unlike the college students—may have known what grade to expect based on their prior performance. Thus, prior experience may have been more important than either their declarative knowledge or their metacognitive knowledge monitoring ability in determining the high school students' estimates.

Summary: Estimates of Performance and Knowledge Monitoring Ability

One striking finding of two of the studies involving college students was that in terms of knowledge monitoring ability, the strongest effects were found for students' actual performance, either on tests or in class, rather than their estimates. Students' estimated performance on the AP exam, or their predicted achievement in class, was typically not significantly related to the KMA scores. On the other hand, performance on the AP test or on final exams (at least the multiple-choice part of the exam in Study VIII) was significantly related to knowledge monitoring ability. These results may be partially attributable to unrealistic estimates of students in the lower knowledge monitoring ability groups.

There was a large difference between the accuracy of vocational high school students and college students in predicting test performance. The correlations between predicted and actual scores for the vocational students were .71 and .75 ($p < .001$) after they took the test; comparable results for college students in Study VII were .13 and .16, both nonsignificant, and for Study VIII they were -.14 and -.12, also nonsignificant. The greater accuracy of the vocational students is probably attributable to their familiarity with the material they were tested on, compared to the unfamiliar content of the AP test for the two college samples. As expected, the relationships between prediction and performance were higher, though not significantly so, after students took the tests, when they knew what was covered.

In both of the studies involving college students, the analysis of academic performance based on actual word knowledge (number correct on the vocabulary test) accounted for more variance than did comparable analyses using the more conclusive KMA scores. It seems possible that students' achievement in class is best predicted by actual word knowledge rather than their estimates of it. Furthermore, in view of the nonsignificant relationships for the high school sample between either actual word knowledge or their KMA scores and final course exam results, it seems likely that domain-specific knowledge may be most useful for predicting course performance.

An important question to investigate is whether estimated knowledge or demonstrated knowledge on tests in the domain which is the subject of instruction and evaluation are likely to account for more variance than similar indices based on fairly general materials, such as those used in the studies reviewed here. The prior research assumed that the word list, vocabulary test, and reading passage were similar to the kinds of material students would be exposed to in nontechnical areas of instruction. The studies relating knowledge monitoring ability to classroom learning found KMA relationships with achievement in English and humanities courses, but not in

the sciences and social sciences. These results suggest that general knowledge, or metacognitive estimates of that knowledge, are less useful in more technical areas that rely on a domain-specific technical vocabulary than they are in subjects that have a more widely shared knowledge base and vocabulary.

Metacognitive Knowledge Monitoring Ability and Mathematics

All of the studies described so far used the KMA in the domain of declarative word knowledge and employed similar or identical versions of the study materials. Therefore, the question arises whether the procedure can be generalized to other domains, such as mathematics. As is true of vocabulary, mathematics is of special interest because it is also important in classroom learning. However, computation and problem solving in mathematics involve more procedural knowledge than does learning vocabulary. Thus, one purpose of the two studies described below was to examine the applicability of the KMA procedure to the domain of procedural knowledge in mathematics.

The research described above involved relatively mature students, predominantly those attending college; only two investigations studied high school students. A further question to be examined in the next two studies was whether the KMA was equally appropriate to examining learning among younger, elementary school students.

*Study IX. Mathematical Problem Solving among Elementary School Students*⁶

Van Haneghan and Baker (1989) reported on a number of investigations of the effects of metacognition on the accuracy of problem representation in mathematics. The results indicated that metacognition was as important for learning mathematics as it was for reading. These findings are supported by other researchers, such as Campione, Brown, and Connell (1989), Lester, Garofalo, and Kroll (1989), and Schoenfeld (1992). Additional research (Cardell-Elawar, 1992; Montague, 1992) has shown that students' performance in solving mathematical problems was facilitated when they were taught a metacognitive approach. Therefore, in the studies below, it was expected that procedural KMA scores in mathematics should be related to overall achievement in that subject.

Participants and Procedures

A set of 30 mathematical questions was constructed

(20 computation and 10 problem-solving items); the items were selected from the fifth-grade mathematics curriculum. Students were first asked to take six minutes to determine whether "you feel able to solve these problems. Do not solve them now," giving them an average of 12 seconds per problem. In a later session, the same 30 questions were used, and students were given 40 minutes to actually solve the problems. A number of anxiety scales were also administered.

A total of 51 fifth-grade students (31 female) from an urban public school served as subjects in this study. The students were predominantly of Hispanic origin and their reading and mathematical achievement ranged from average for their grade to two years below grade level.

Results and Discussion

Scoring of the mathematics KMA was similar to the procedure used for the word knowledge KMA scores reported earlier and involved generating four scores: Students felt that they could:

- 1) solve a problem and did so (++)
- 2) not solve a problem and did not (--)
- 3) solve a problem, but did not (+-), and
- 4) not solve a problem, but did (-+).

The results dealing with anxiety will be discussed later in this report.

There were no differences among students' metacognitive estimates of ability attributable to gender, so these data were pooled for further analysis. The knowledge monitoring scores were correlated with the total math score on the Metropolitan Achievement Test (1985) obtained from the students' records. The correlations are displayed in Table 4. The last row in that table represents the number correct on the math test. The ++ and -- scores were combined to indicate students' correct estimates of their

Table 4

Correlations Between Knowledge Monitoring Scores and Achievement in Mathematics*

<i>Knowledge Monitoring Score</i>	<i>Correlation</i>
++	.73***
+-	-.43**
-+	-.65***
--	-.11
++ and --	-.76***
-+ and +-	-.72***

* The correlation between the Knowledge Monitoring raw score (i.e. total number correct) and performance on the Metropolitan Achievement Test in Math was .52**.

** $p < .01$
*** $p < .001$

ability to solve mathematical problems, and the $-+$ and $+-$ scores were added to show the incorrect estimates.

Table 4 indicates that three of the four estimates were significantly related to students' achievement in mathematics. The correlation between the number correct on the math test and the Metropolitan test score was .52. When that relationship is compared to the correlation of .73 between the Metropolitan test score and $++$, or the correlation of .76 between the Metropolitan test score and total number of correct estimates, it is clear that metacognitive estimates of ability to answer the questions are more substantially related to mathematical achievement than is the number of problems solved correctly, irrespective of estimate. That finding was confirmed by regression analysis. When the number of correct estimates, incorrect estimates, and total number right were used in the regression, only the correct estimates contributed significantly to prediction of the Metropolitan test score [R^2 Change = .08, $F(3,45) = 8.52$, $p < .01$]. These results confirm the basic assumption that students' metacognitive estimates of their ability contribute significant independent variance beyond that accounted for by the number correct on a test.

The results support predictions regarding the relationships between the procedural KMA in mathematics and achievement in that domain. As expected, there were substantial correlations between students' estimates of their ability to solve mathematical problems and their achievement in mathematics. Also as expected, inaccurate estimates were negatively related to achievement. While no causal inferences about the relationship between mathematical achievement and the ability to monitor knowledge can be made from these correlational data, the fact that the variables co-vary as expected supports the generalizability of the KMA procedure and suggests that the technique is useful for further research on achievement in mathematics.

Study X. Relationship of Knowledge Monitoring Ability in Mathematics to Age and Achievement⁷

The prior study provided encouraging evidence of the applicability of knowledge monitoring ability to achievement in mathematics. Furthermore, the results of Study IX also indicated that the KMA could be used in assessing elementary school students. Because metacognition is often viewed as a developed ability and assumed to increase with age, one purpose of this study was to investigate whether procedural knowledge monitoring ability in mathematics would also increase with age. The preceding study indicated a high relationship between KMA scores in mathematics and achievement test scores in that domain. Study X examined whether knowledge monitoring scores were related to teachers' judgments of mathematical ability.

Participants and Procedures

Students ($N = 164$, 70 female) were selected from the fourth, fifth, and sixth grades of a school attended largely by minority students. Mathematical ability was determined by teachers' judgments; 29 students were placed in the low-, 93 in the medium-, and 42 in the high-ability groups.

Students were presented with 15 mathematical word problems involving addition and subtraction. The problems were set in the context of an ice cream store and students received a menu of prices for different products that were referred to in the problems. The materials were prepared in two versions presumed to elicit varying levels of interest among students. The results dealing with interest will be discussed later in this report. The materials were administered on two days during regular class periods. On the first day, students examined the problems and estimated whether they could solve them or not; on the second day, the students were asked to solve the problems.

Results and Discussion

Students' responses were assigned a score of 1 for each correct estimate (combining the $++$ and $--$ scores) and 0 for each incorrect estimate (combining the $+-$ and $-+$ scores). Due to a computer malfunction, raw data were not available for rescoring in the format used in the other studies. The data were then submitted to a 3 (grades) \times 2 (gender) \times 2 (group: control versus interest, see below) \times 3 (math ability) analysis of variance.

As expected, there was a significant increase in knowledge monitoring scores from grades four to six ($F = 34.66$, $df = 2, 144$, $p < .001$, $\eta^2 = .26$; see Figure 7 for a plot of the data). Also as expected, knowledge monitoring scores increased with mathematical ability ($F = 15.25$, $df = 2, 144$,

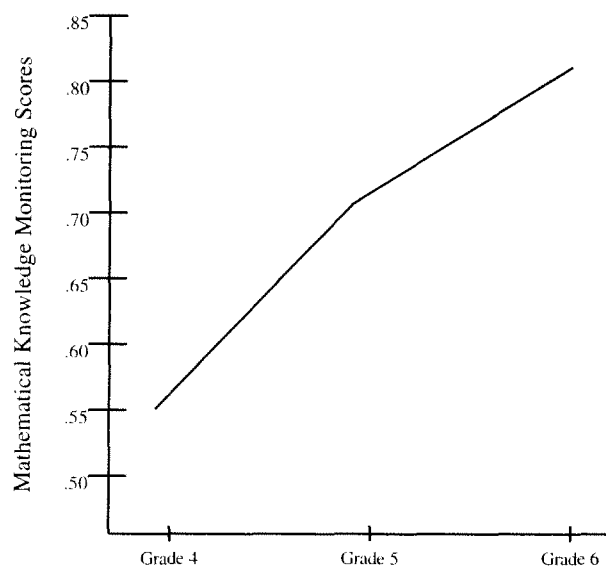


Figure 7. Relationships between mathematical knowledge monitoring scores and grade level.

$p < .001$, $\eta^2 = .18$; see Figure 8 for a plot of the data). These results offer further support for the construct validity of the KMA procedure in that older or more capable students were expected to have higher knowledge monitoring ability than their younger, less capable counterparts. There were no significant differences attributable to gender.

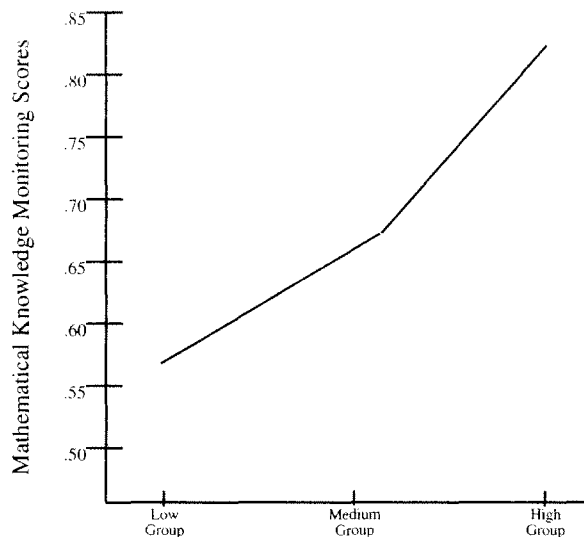


Figure 8. Relationships between mathematical knowledge monitoring scores and math achievement group.

Summary: Knowledge Monitoring Ability and Achievement in Mathematics

The results of Studies IX and X were quite positive regarding the applicability of the KMA to mathematics. The relationship of knowledge monitoring scores to achievement in mathematics in Study X is similar to the correlations with math achievement test scores reported in Study IX and both indicate strong relationships between metacognitive knowledge monitoring ability and achievement in mathematics. The increases in ability associated with age reported in Study X also support that relationship. Furthermore, because most of the items used in both mathematics studies involved procedural knowledge of the type needed to solve word problems, the results suggest that the KMA may be applicable to procedural knowledge as well as declarative word knowledge.

Metacognitive Knowledge Monitoring Ability and Affect

The paradigm shift to a cognitive orientation in psychology generated a great deal of research intended to clarify the cognitive processes controlling learning. However, the impact of affective processes on learning has received considerably less attention (Tobias, 1992, 1994a,

b). The research discussed in this section was intended to forge a link between affect and cognition by examining the influence of affective variables such as anxiety and interest on metacognitive knowledge monitoring ability.

The Impact of Anxiety on Knowledge Monitoring Ability

One affective variable that has been the subject of a great deal of research is anxiety and its impact on learning. In general, that research has suggested a negative relationship between different forms of anxiety and achievement (Tobias, 1992, Hembree, 1988). It has been suggested (Tobias 1985, 1992) that anxiety reduces the cognitive capacity available for task solution. The capacity required by an executive process such as metacognitive knowledge monitoring was expected to be especially reduced among highly anxious students. Therefore, a negative relationship between anxiety and knowledge monitoring ability was anticipated because “highly test anxious students can be expected to have less adequate metacognitive abilities than those with lower anxiety” (Tobias, 1992, p. 28).

Knowledge Monitoring Ability, Reading Comprehension, and Test Anxiety

Study II also examined the relationship of the KMA procedure to anxiety. The worry subscale of the Test Anxiety Inventory (Spielberger, et al., 1980) was administered to the subjects.

As expected, the more highly anxious participants performed less well on the KMA. Those with less anxiety achieved a significantly higher number of “hits” than those prone to higher levels of anxiety [$t(115) = 4.92$, $p < .001$], and in general the less anxious subjects had higher levels of metacognitive word knowledge as measured by d' , [$t(115) = 4.07$, $p < .001$], confirming the expected negative relationships between knowledge monitoring ability and test anxiety.

Knowledge Monitoring Ability in Mathematics and Anxiety

Study II found the expected negative relationship between knowledge monitoring scores and anxiety with respect to vocabulary. Study IX, in addition to investigating the extension of the KMA procedure to mathematics, also studied its relationship to both test and mathematics anxiety.

As part of Study IX, the Fenema-Sherman (1976) scales assessing math anxiety and attitudes toward mathematics were administered to the participants (see the earlier description of Study IX) in a first session. To ensure that the

subjects could understand the questions, each item was read aloud while the students read the materials to themselves. The Worry-Emotionality Scale (Morris, Davis, & Hutchings, 1981), a 10-item, Likert-type measure of these components of test anxiety, was also administered. Students' achievement in mathematics was determined from their scores on the Metropolitan Achievement Test (1985) obtained from school files.

In Study IX no gender differences in the effects of anxiety were found, so the data for all students were pooled. The relationships between knowledge monitoring ability and mathematics anxiety, as well as with worry and emotionality, are shown in Table 5.

Table 5

Correlations Between Knowledge Monitoring Scores and Anxiety in Mathematics

<i>Score</i>	<i>Math Anxiety</i>	<i>Worry and Emotionality</i>
++	-.42**	-.22
+-	.32*	.25
-+	.38**	.23
--	.00	.20
++ and --	-.46**	-.15
-+ and +-	.46**	-.33*

* $p < .05$

** $p < .01$

Table 5 indicates that, as expected, mathematics anxiety was negatively related to incorrect estimates of knowledge and positively related to correct ones. The negative relationships between knowledge monitoring ability and anxiety are generally similar to those found in Study II, confirming expectations that anxious students have less ability to monitor their knowledge than their less anxious peers.

Knowledge Monitoring Ability and Anxiety among High School Dropouts and Continuing Students

Study VI investigated whether continuing students and high school dropouts differed in knowledge monitoring ability. An additional purpose of that study was to examine the differences in anxiety level between continuing students and high school dropouts, as well as the relationship between anxiety and metacognitive knowledge monitoring. In this study, the Test Anxiety Inventory (Spielberger et al., 1980) was given to all participants, followed by two administrations of the Worry-Emotionality Scale (Morris et al., 1981). Initially, par-

ticipants were asked to complete the Worry-Emotionality Scale in terms of how they felt while being tested in general; when the scale was readministered after the vocabulary test, students were asked to complete the scale in terms of how they felt while completing the vocabulary test.

Surprisingly, the results of a MANOVA indicated that there were no differences in anxiety level between high school dropouts and continuing students on any of the seven anxiety scores (the three Test Anxiety Inventory scores: Worry, Emotionality, and Total, in addition to four Worry and Emotionality scores from each administration of those scales). That finding is puzzling in view of the reports in the literature that poor performance in school, and presumably on tests, is a major reason students drop out of high school. One explanation may lie in the problems to which self-report measures in general, and self-reports of test anxiety in particular, are subject. Students can easily minimize or deny indications of test anxiety when responding to these measures and present themselves as not caring about how well they might function on tests. The KMA procedure, however, made it difficult for students to present themselves in a more favorable light, and that may account for the findings of group differences in metacognitive knowledge monitoring ability and the absence of differences on measures of test anxiety.

Most of the zero-order correlations between the KMA scores and the anxiety indices were negative, and a fair number were significant. Multiple linear regression analyses were computed with the KMA scores as the dependent variable and the anxiety scores as the independent variable. Results indicated that the anxiety scales had a significant impact only on the ++ scores [$R^2 = .25$, ($F(7,72) = 3.43$, $p < .01$); significant beta weights were found for Emotionality on the Worry-Emotionality Scale taken after students had completed the vocabulary test ($t = 2.74$). The regression analysis also indicated that none of the other KMA scores was significantly related to the anxiety scales. In view of the number of anxiety and knowledge monitoring scores, the finding of significant relationships for some of them is not surprising. In general, however, the results of this study suggested that there was little association between metacognitive knowledge monitoring ability and anxiety.

Knowledge Monitoring Ability and Anxiety among Vocational High School Students

Study V, which examined knowledge monitoring ability among vocational high school students, also investigated the relationship between anxiety and knowledge monitoring ability, as well as between anxiety and achievement. In addition to relating metacognition to anx-

ity, it was expected that students with lower GPAs would have higher levels of anxiety than those with higher GPAs.

The anxiety scales and the order in which they were administered in Study V were identical to those used in the study of high school dropouts (Study VI). The Worry-Emotionality Scale (Morris et al., 1981) was administered first and students were asked to respond in terms of how they felt while taking tests in general. The Test Anxiety Inventory (Spielberger et al., 1980) was then given, followed by a second administration of the Worry-Emotionality Scale with instructions for students to respond in terms of how they felt while taking the vocabulary test.

The significance of the differences in anxiety scores between the participants in Study V above and below the median GPA was examined by performing a MANOVA. Surprisingly, there were no differences between the two GPA groups on any of the seven anxiety scores. Also, much as in Study VI, most of the zero-order correlations between knowledge monitoring ability and level of anxiety were negative. Multiple linear regression analyses were then computed with the knowledge monitoring scores as the independent variable and the anxiety scores as the dependent variable. None of the regression equations was significant for this sample.

Summary: Knowledge Monitoring Ability and Anxiety

The evidence regarding the relationship between level of anxiety and knowledge monitoring ability is mixed. Significant negative relationships were expected and found in two of the studies, one in mathematics and the other using vocabulary materials. On the other hand, two additional studies failed to find any evidence of differences. There was a larger sample in Study II, which found significant negative relationships with level of anxiety using the vocabulary materials than in the studies involving vocational high school (Study V) or high school dropout (Study VI) groups. Because many of the test anxiety–metacognitive knowledge monitoring relationships in the latter two studies were, as expected, in the negative direction, and because some of the regression analyses approached significance, further research with larger samples is needed to clarify the relationship between anxiety and knowledge monitoring ability. The results of Study II suggest that knowledge monitoring ability and level of test anxiety each contributed to performance on less challenging reading material. On more demanding material, however, test anxiety and knowledge monitoring ability appeared to interact to affect performance. The highly anxious examinee, regardless of metacognitive ability, performed less well on the more demanding reading tasks, suggesting that worrying can interfere with strategic use of

metacognitive skills when tasks are cognitively demanding. This finding is in accord with the anxiety-cognitive capacity model (Tobias, 1992) in that more demanding tasks require greater cognitive capacity that may not be available because of the resources absorbed by anxiety. Further research is required to pursue that intriguing finding.

In Studies V and VI, the failure of a number of anxiety indices to differentiate between either high school dropouts and continuing students, or between students above and below the median in GPA, was surprising. A meta-analysis of 562 studies dealing with test anxiety (Hembree, 1988) indicated that lower-achieving students experienced more test anxiety than did their more capable counterparts. While there had been no prior research specifically relating test anxiety to dropping out of high school, the bulk of the literature has indicated that students' concern about their academic achievement was a major factor in dropping out of school, clearly suggesting that differences in test anxiety could be expected. As mentioned above, the fact that the studies dealing with dropouts and vocational high school students both found significant differences in knowledge monitoring ability but neither found differences on a group of seven test-anxiety scales reemphasizes some of the problems with self-report measures, described at the beginning of this report.

While the nonsignificant results for anxiety in Studies V and VI may be attributable to the small samples, or to other unknown factors, it should also be noted that the tendency of participants to present themselves in a more positive light may well have contributed to the nonsignificant findings. One advantage of the KMA is that, because students do not report on either their feelings or their cognitive processes, it is difficult for them to present themselves more favorably. Of course, students could easily claim to know more words or to be able to solve more problems than is actually the case. However, that claim will be immediately challenged by administration of the test, making it harder for students to appear in a more positive light.

Knowledge Monitoring Ability and Interest

A good deal of recent research has addressed the effects of interest on learning for a variety of reasons (Renninger, Hidi, & Krapp, 1992). Clarification of the effects of interest adds to an understanding of the impact of intrinsic motivation on learning. Interests also appear to be stable and long-lasting among adults (Hidi, 1990; Schiefele, 1991), suggesting that instruction adapted to students' interests may have positive motivational effects over long periods of time. In addition, interests are ubiquitous in that everyone is interested in something. Also, findings of surprisingly

variable and ineffective cognitive processing (Paris, 1988; Tobias 1989) suggested that students' interests or motivation may not have been engaged by the materials used in the studies. Finally, research on interest provides a useful and educationally relevant avenue for studying the relationship between affect and cognition (Tobias, 1989, 1994a, b), which is needed to obtain a more complete picture of how individuals function on a day-to-day basis.

Research has indicated that reading comprehension and recall are facilitated when students work on material related to their interests (Renninger et al., 1992). Furthermore, Schiefele (1990, 1991, 1992a, b) found that comprehension of interesting text was "deeper" (i.e., more likely to be propositional than verbatim). Little is known, however, about the cognitive processes that mediate the effect of interest on comprehension and recall. Therefore, it was recommended (Renninger et al., 1992; Tobias, 1994a) that research concentrate on the identification of the processes invoked by interest that facilitates learning. The studies reported in this section examined whether interest improved students' metacognitive knowledge monitoring ability.

*Knowledge Monitoring Ability, Situational Interest, and Topical Interest*⁸

Two types of interest, situational and topical, have been distinguished (Renninger et al., 1992). Situational interest is elicited by aspects of a situation, such as its novelty or intensity, and by the presence of factors contributing to the attractiveness of different types of content. Topical interest refers to individuals' relatively enduring preferences for certain topics, tasks, or contexts and how these influence learning. The effects of both types of interest on knowledge monitoring ability were investigated in this study. It was expected that subjects with greater topic interest to students and text that elicited situational interest would generate more accurate knowledge monitoring. Furthermore, because interest was found to lead to deeper text processing (Schiefele, 1990, 1991, 1992a, b), it was expected that students would make more accurate knowledge monitoring estimates on words requiring intense processing if the material were interesting to them. In addition, because meanings of implicitly defined words have to be inferred, whereas those defined explicitly merely require recall of the definitions, it was reasoned that the meanings of implicitly defined words should be estimated more accurately when content was interesting to subjects.

Study III Revisited

It will be recalled that there were two groups of students in Study III, nursing students and college freshmen. Because the reading passage dealt with heart disease, it

was expected that nursing students would have a greater topical interest in that material than would freshmen. Situational interest was varied by converting the expository passage to a narrative format. The narrative passage contained story attributes, such as character identification and life themes, which, according to Hidi and Anderson (1992), should increase situational interest. A principal character was introduced in the narrative version, which then described his efforts to learn more about coronary disease because his father had developed a mild form of that illness. The passage indicated that he was trying to help his father prevent more serious coronary problems. This structure made it possible to include all the factual information presented in the expository version of the passage. Of the 139 students in the study, 84 completed all the materials during the two sessions. Complete data were available for 33 nursing students and 51 freshmen.

In Study III, an analysis of variance was performed on the correct metacognitive estimates (combining ++ and -- scores), with the correct estimates on explicitly and implicitly defined words—the dependent variables—treated as a repeated measure. In view of the importance of controlling for differences in prior knowledge (Tobias, 1994), students' scores on the first administration of the vocabulary test were used as a covariate because the nursing students were more familiar with the heart disease material (prescore mean = 27.4, S.D. = 4.0) than the freshmen (prescore mean = 20.1, S.D. = 5.3). Because there was an unequal number of females in the groups (24 of 51 freshmen and 28 of 33 nursing students), gender was added as a factor. Thus, the ANOVA consisted of a full 2 (freshmen vs. nursing students) x 2 (expository vs. narrative passages) x 2 (gender) factorial design, with prescore as a covariate. Again, the two-level repeated measure consisted of the number of correct estimates on explicitly and implicitly defined words after subjects read the passage. The main effect of the repeated measure was assessed in the "deviation" manner (Delaney and Maxwell, 1981).

The ANOVA results indicated that there was a significant overall difference between the freshmen and nursing students [$F(1,75) = 4.99, p < .05$] favoring the nursing students. In addition, the mean number of correct estimates was higher for explicitly than for implicitly defined words [$F(1,75) = 8.27, p < .01$]. None of the other main effects or interactions was significant. The covariate, the number correct on the first administration of the vocabulary test, exerted a significant effect on the dependent measures [$F(1,75) = 17.01, p < .001$]. The adjusted means for freshmen on correct estimates for explicitly and implicitly defined words were 13.7 and 12.5, respectively, and for nursing students the corresponding means were 15.0 and 14.1.

These results support the general hypothesis that topical interest enhances metacognitive knowledge monitoring ability. As anticipated, nursing students, for whom the heart disease passage was more interesting, made more accurate metacognitive estimates of their vocabulary knowledge than did the freshmen, even when differences in prior knowledge of the vocabulary were controlled for statistically. The expected differences attributable to situational interest were not found because the KMA scores for the narrative and expository passages were similar. Finally, contrary to expectations for both nursing students and freshmen, explicitly defined words were estimated more accurately than those that were implicitly defined.

The absence of differences in knowledge monitoring ability due to situational interest may be a function of the similarities between the expository and narrative texts. Even though the passage was altered to create differences in situational interest, ratings of interest on a Likert-type scale, in the original study and on a follow-up, failed to show any differences between the passages. Perhaps greater differences in content are needed to produce differences in situational interest.

Knowledge Monitoring Ability and Interest in Mathematics among Elementary School Students

Study X found that metacognitive knowledge monitoring ability in mathematics increased with grade and mathematical ability. A further purpose of that study was to examine the impact of personalizing instruction on metacognition. Research (Anand & Ross, 1987; Bracken, 1982; Herndon, 1987; Lopez, 1990, 1989; Ross & Anand, 1987; Wright & Wright, 1986) has shown that personalizing mathematical word problems by including materials such as the names of students, their friends, or teachers, or including materials related to students' interests improved performance and attitudes toward the materials. It was, therefore, hypothesized that increased interest generated by personalizing the word problems should improve students' knowledge monitoring ability.

Participants in Study X were randomly assigned to either interesting or control materials. In the "interesting" materials, the names of classmates and teachers were included in the math word problems, whereas the materials used for the control group contained standard names. In each set of materials, 15 mathematical word problems, set in the context of an ice cream store, were presented. Students received a menu of prices for different products and were required to add and subtract prices of menu items. A 12-item Likert scale designed to

assess interest in the materials was also administered.

In this study, students' responses were assigned a score of 1 for each correct estimate and 0 for each incorrect estimate of their knowledge. The data were then submitted to a 3 (grades) \times 2 (gender) \times 2 (interesting or control materials) \times 3 (math ability) analysis of variance. The findings dealing with knowledge monitoring ability, mathematical achievement, and grade level were reported previously. There were no significant differences attributable to gender or to interest. However, there was an interaction between math achievement level, as determined by teachers' evaluations, and interest ($F = 6.02$, $df = 2.144$, $p < .01$, $\eta^2 = .05$; see Figure 9 for a plot of the data).

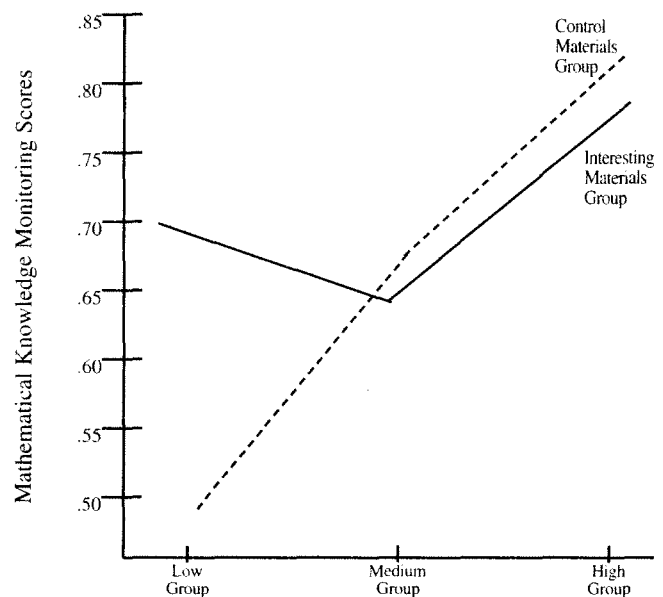


Figure 9. Relationships between mathematical knowledge monitoring scores, interesting or control materials group, and math achievement group.

The interaction, unlike the main effect found in the previous interest study, suggests that personalization improved the performance of low-ability math students but had little effect on the two other groups. In view of the known difficulties students have with math word problems (NAEP, 1979), it was thought to be important to make the materials interesting for both groups by creating an ice cream parlor setting. It seems possible that setting the math word problems in this context may have made the materials more interesting for both groups, thus leading to the insignificant main effect for interest. There is evidence that this setting did arouse the interest of all students. There were no differences ($F < 1.0$) between the high- and low-interest groups on the 12-item Likert scale administered after students completed the problems. Furthermore, there were no differences between the high- and low-interest groups in the number of problems solved correctly. These findings indicate that

even the low-interest group may have found the materials more attractive than the math word problems usually presented in class and suggest that an overall effect might be found if there were greater differences in interest between the groups.

Summary: Knowledge Monitoring Ability and Affect

The findings of the anxiety and interest studies indicate that anxiety generally seems to have a negative effect on metacognitive knowledge monitoring ability and that working on interesting materials seems to facilitate it. Further research is needed to answer many questions before these tentative conclusions can be stated with greater confidence. It seems, however, that the KMA procedure is a useful tool for studying the effects of affect, and especially of interest, on metacognition. There are a number of persuasive models specifying the cognitive processes mediating the impact of anxiety on learning (Sarason, 1987; Eysenck, 1988; Tobias, 1992). However, little is known about the cognitive processes by which such "positive" affective variables as interest and motivation facilitate learning. The KMA procedure seems to be of use for further research relating metacognition to such positive variables as interest or intrinsic motivation.

Metacognitive Knowledge Monitoring Ability and Other Variables

Two additional studies examined the relationship of the KMA procedure to the need for feedback and the KMA's ability to differentiate between different types of students. These are summarized below.

Study XI. Knowledge Monitoring Ability and Need for Feedback⁹

Feedback or reinforcement is one of the most widely studied variables in learning research. Numerous studies have demonstrated that feedback facilitates learning. McKeachie (1974) suggested that the effects of feedback or reinforcement on learning are not uniform but may vary with individuals and situations. Ashford and Cummings (1983) found that the importance of feedback varies with an individual's uncertainty and Tuckman and Sexton (1992) found that students in a no-feedback situation who held high expectations for their own performance outperformed those receiving feedback, whereas the reverse was true for students of middle and low self-

perceived ability. These results clearly support the idea that there are individual differences in the need for feedback.

It was expected that the need for feedback would depend on students' metacognitive ability to monitor their knowledge-gathering activities. In an analysis similar to that proposed by Butler and Winne (1995), it was proposed that students with accurate knowledge monitoring ability probably rely more frequently on their own internal feedback regarding the accuracy of their responses than do their less-accurate peers. Such students are likely to have learned from experience that external feedback often duplicates the information supplied internally and they therefore should require less externally supplied feedback than do peers with less accurate knowledge monitoring ability. Therefore, when students had a choice of whether to obtain feedback or not, a negative relationship between KMA scores and amount of feedback was expected.

Participants and Procedures

A sample of 59 fifth-grade students (35 female) from a predominantly minority school participated in this study. A list of 25 words appropriate for fifth-grade students and a vocabulary test based on the same words were developed. Participants were also given a reading test consisting of 11 narrative stories with an average length of 140 words or 15 sentences. Each story had a blank to be filled in, and students were instructed to select a word for each blank from four choices appearing in the right margin. The words on the word list and reading test were different. Participants were told that the correct answer to each question was printed in the left margin of each page, covered by a tab, and that they could look at the answers whenever they wished to by simply lifting the tab. Participants were tested individually, and the number of times the tabs were lifted to check the correct answer was recorded.

Results and Discussion

Students' need for feedback was operationally defined as the number of times they lifted the tabs covering the correct answers. The KMA procedure was used to determine students' accuracy in estimating their word knowledge, and the results were then correlated with amount of feedback sought. The results of that analysis are shown in Table 6.

As expected, the results indicate that amount of feedback needed is substantially related to students' ability to accurately monitor their knowledge. Accuracy of knowledge monitoring was substantially and negatively related to amount of feedback ($r = -.79, p < .001$), as was the number of inaccurate estimates ($r = .76, p < .001$). Equally interesting was the finding that vocabulary knowledge,

Table 6

Correlations Between Knowledge Monitoring Scores and Need for Feedback

<i>Knowledge Monitoring Score</i>	<i>r</i>
++	-.50**
+-	.38*
-+	.56**
--	-.13
++ and --	-.79**
-+ and +-	.76**
Actual score (total correct regardless of estimate)	-.19
<i>r</i>	.84**
<i>r</i> ²	.71

* $p < .01$

** $p < .001$

determined by the number correct on the vocabulary test, was not significantly related to amount of feedback ($r = -.19$). The findings suggest that, as expected, students' need for feedback is strongly related to their ability to accurately monitor their knowledge. Furthermore, students' estimates of their knowledge were clearly the major contributor to that relationship given that actual knowledge was unrelated to amount of feedback.

An equally important aspect of this study and its results was the fact that a new word list and vocabulary test were developed, different from the materials used in any of the other studies described in this report. Therefore, the findings also indicated that the KMA procedure has some generality across different types of vocabulary materials. Furthermore, this was the first study using a declarative vocabulary KMA with elementary school students, and the results suggest that the procedure is as applicable to younger students as were the mathematical materials used in Studies IX and X.

Study XII. Differences in Knowledge Monitoring Ability among Learning-Disabled and Hyperactive Students¹⁰

It has been shown (Brown & Campione, 1986; Swanson & Trahan, 1992) that students diagnosed as Learning Disabled (LD) have lower metacognitive monitoring ability than do those without special needs. Students with Attention Deficit Hyperactivity Disorders (ADHD) have been succinctly described by Douglas, Barr, O'Neil, and Britton (1986) as having an inability to stop, look, listen, and think, which also has a negative effect on metacognition. A review of research dealing with ADHD (Westby & Cutler, 1994) indicates that such students tend

to have less effective complex problem-solving strategies and organizational skills, that they use less efficient strategies on memory tasks, that they "demonstrated deficits on all measures of study behavior. They studied for less time, expended less effort, and used poorer strategies...students with ADHD have significant deficits in executive processes." (Westby & Cutler, 1994, pp. 63-64.) These deficits clearly suggest that ADHD students have less effective metacognitive ability. Therefore, students diagnosed as LD or ADHD should have less accurate knowledge monitoring ability than students not affected by these conditions. This study tested that hypothesis.

Participants and Procedures

A list of 35 words and a vocabulary test based on the same words were developed from the high school curriculum. Participants ($N = 90$) were selected from the ninth ($N = 29$) and tenth ($N = 61$) grades of a public high school in an urban area; there were 28 females and 62 males. LD and ADHD groups ($N = 30$ each) were formed by selecting students diagnosed by a school-based support team consisting of a licensed educational evaluator, a school psychologist, and a social worker. Scores on the Degrees of Reading Power (DRP) (Touchstone, 1991) test placed these groups in the fifteenth percentile of the population. A control student group ($N = 30$) was selected on the basis of demonstrating average reading ability on the DRP and having no history of special educational needs.

Results and Discussion

Three of the KMA scores (++, +-, and --) were analyzed using MANOVA (the fourth score, -+, could not be entered due to linear dependencies), with gender and group as the independent variables. A significant overall difference among the groups was found [Wilks $F(6,164) = 5.95$, $p < .001$, effect size = .179]. Univariate analyses indicated significant differences between the groups on ++ scores [$F(2,84) = 16.02$, $p < .001$; control group mean = 28.4; LD mean = 22.2; and ADHD mean = 23.0]. Univariate analyses also indicated another difference on the -- score [$F(2,84) = 5.32$, $p < .01$; control group mean = 1.5; LD mean = 3.6; and ADHD mean = 4.3]; students in the control group had lower scores because they had fewer incorrect answers. There were no differences attributable to gender, and no interaction between gender and group was found.

A similar analysis of the number correct on the vocabulary test (++ and -+) also indicated significant group differences [$F(4,166) = 7.55$, $p < .001$, effect size = .154]. Univariate analysis indicated that only the differences on the ++ scores were significant; the group means were the same as for the preceding analysis. The results confirm expectations regarding differences between regular, LD, and ADHD students with respect to their ability to mon-

itor their knowledge and differentiate between what they know and do not know in this domain. While the results were similar when the dependent variable consisted only of the number correct on the vocabulary test, the effect was smaller (.154 compared to .179). As expected, the control group of students without special needs received KMA scores showing that they were better able to differentiate between the ++ and -- words than were students in the other two groups.

There were large differences in reading ability between the groups, which may also have accounted for the group differences, irrespective of diagnostic category. It is often difficult to separate the effects of reading ability in research comparing LD, ADHD, and more traditional students because reading problems are one of the defining characteristics of the two former groups. Further research with similar groups may resolve this problem. In any event, these results provide additional support for the construct validity of the KMA procedure. In view of the fact that this study, like the prior one, also developed a new list of words and vocabulary test, the results also support the generality of the KMA procedure across different types of vocabulary materials.

General Discussion

The findings of the 12 studies summarized above support the construct validity of the KMA procedure. Comparable results were found for samples from diverse student populations: elementary school students, those attending regular and vocational high schools (including students diagnosed as LD and ADHD), those who dropped out of high school, students in pre-freshman skills programs, and those who had attended college for some time. Furthermore, substantially similar results were obtained for procedural knowledge in mathematics, as well as for declarative vocabulary knowledge based on three different sets of vocabulary materials developed to be appropriate for students at elementary school through college levels.

In view of the fact that the KMA may be administered to groups, as well as to individuals by computer, and is objectively scored, it seems to be a promising approach for the assessment of the knowledge monitoring component of metacognition. In addition, Studies V and VI indicated that the KMA made it less likely that students could present themselves in a more favorable light, one of the problems inherent in the social desirability aspect of self-report instruments. While no data comparing the KMA to other metacognitive scales have so far been collected, we expect that this measure of knowledge monitoring ability is likely to be more accurate than self-report scales because students are less able to present

themselves in an artificially favorable way. It remains for further research to investigate this correlation.

KMA relationships with external criteria were somewhat variable. Relationships with standardized achievement tests were substantial and significant. For example, in Study I, correlations with a reading comprehension test were .67. Relationships with achievement in mathematics were also substantial in Study IX ($r = .76$); and in Study X, highly significant effects were found for KMA differences in students' math achievement ($\eta^2 = .26$) and for increases in mathematical ability over these elementary school grades ($\eta^2 = .18$). Pintrich (in press) cites some of these findings as being among the most positive relationships identified between any metacognitive measure and external criteria. Relationships with need for feedback (Study XI) were also found to be substantial ($r = .62$). Significant, though somewhat more moderate, relationships were found in those studies in which the KMA differentiated between divergent groups such as regular students and dropouts (Study VI), or among LD, ADHD, and students without special needs (Study XII). Generally, the lowest, though frequently significant, relationships were found between KMA scores and college grades. Presumably, as indicated previously, the low reliability of such grades accounts for the modest effects. Further, differences between the effects of knowledge monitoring estimates and actual knowledge discussed below should also be considered.

A number of issues raised by the results require further research. These include the following: Do multiple administrations of the KMA procedure strengthen its relationship with other variables? Which of the different scores are optimal indicators of knowledge monitoring ability? Do estimates of knowledge account for more variance than actual knowledge? These questions are addressed below.

The KMA Procedure and Dynamic Assessment

Some of the studies described above administered the reading passage to only a portion of the sample, others did not use the passage at all, and still others administered a word list and vocabulary test before and after students read a passage from which the word meanings could be inferred. A question arises about the value of interjecting the reading passage between administrations of the word list and vocabulary test. Giving students a chance to improve their knowledge has some similarities to dynamic assessment approaches (see Carlson & Wiedl, 1992; Guthke, 1992; Lidz, 1992), in which students are given new learning opportunities before being tested. Dynamic assessment procedures usually also include as part of the assessment process some intervention in students' attempts to learn, observations of their reactions to the

intervention, and an evaluation of their responses to the assistance. Reviews have suggested (Carlson & Wiedl, 1992) that students' attempts to verbalize their learning difficulties, and receiving elaborate feedback about their efforts, contribute heavily to the value of dynamic assessment. The KMA differs from dynamic assessment procedures because it does not include any of these additional efforts to facilitate learning; students are merely given a second opportunity to learn the words by reading a passage, without any other assistance.

The results of the present research indicate that the opportunity to learn the meanings of some words from the passage was most important only in Study I, relating the KMA procedure to reading comprehension, and seemed to have little effect on studies of learning in college or estimation of actual performance. The findings indicate that, with the possible exception of relationships with reading comprehension, use of the word list and vocabulary test alone appear to be effective in estimating metacognitive knowledge monitoring ability, whether the reading passage is used or not.

The distinction between explicitly and implicitly defined words was expected to be useful only in those studies in which students read the passage. The results of those investigations indicated that there were few differences between these two types of words. Since neither the use of the passage nor the distinction between the two types of words appeared to affect the results, it seems prudent to abandon both those approaches in future research.

Implications for Teaching and Research

The results indicated that use of the reading passage did not add much explanatory power to the KMA as an appraisal instrument. It may, nevertheless, be interesting to use the passage in future research to study the applicability of the KMA for research on teaching students how to monitor their own knowledge. If the word list and vocabulary test are used as pre-post measures, the passage could be interjected to help students learn the meanings of those words about which they had made incorrect estimates of knowledge. Different levels of instructional support (Tobias, 1989) could be used to help students learn the meanings of the words they had estimated incorrectly.

Use of the reading passage makes it possible to implement a teaching strategy featuring maximal prompting, in the form of very active instructional interventions at the beginning and fading that out until the passage alone is presented without any prompts. The interventions could include such procedures as urging students to provide definitions or synonyms for the words, asking them to rephrase the clauses containing the target words, asking questions about the words and cuing students that the

target words are especially important and that they should pay special attention to them. Of course, research would have to determine whether the suggested interventions actually constitute a hierarchy ranging from maximal to minimal instructional support. It should also be noted that a number of passages, with associated word lists and vocabulary tests, may be needed to develop effective strategies for teaching students how to monitor their own knowledge. Once research has determined the usefulness of the procedures outlined above, they could become an important resource to help teachers at all levels improve the knowledge monitoring ability of their students.

In addition to the possible value for teaching of the instructional interventions described above, interjecting reading passages or content material could make the KMA more similar to dynamic types of assessment and to actual classroom learning. Research could then determine whether such interventions improve the relationship of the KMA procedure to classroom learning. Giving students new learning opportunities before administering, or re-administering, the KMA procedure is likely to be more complex in mathematics or science than for word knowledge. Dynamic assessment in these fields would probably require very active instructional interventions before students show increases in knowledge, because few learners can master new material in science or mathematics merely by reading a passage and working on problems, or even when assisted by the types of interventions suggested above.

Optimal Indicators of the Latent Knowledge Monitoring Construct

Metacognitive knowledge monitoring ability is a latent construct inferred from the various scores generated by the KMA procedure. Many of the preceding studies combined the ++ and -- scores to develop a measure of knowledge monitoring ability. The combined score seemed to have face validity as the most direct and most theoretically interesting index of knowledge monitoring ability. Furthermore, by including the -- scores, the combined total seemed independent of students' actual knowledge, because the combined estimate included items answered incorrectly. Scores based on the signal detection paradigm were used in Study II, but seemed to add little to the combination of ++ and -- scores used in the other studies. However, the findings of some of the investigations, especially Studies VII and VIII, suggest that differences between groups were obscured when the subscores for different categories (++ , +- , -+ , and -- for words defined explicitly or implicitly) were combined.

Ideally, the optimal KMA score should be determined empirically, rather than on the basis of its face validity. The four subscores, or eight if the explicit-implicit distinc-

tion is used, should be submitted to procedures such as the analysis of covariance matrices in order to determine which score(s) are optimal indicators of the latent knowledge monitoring construct. Further research is clearly needed with larger samples (perhaps 200 to 300 students) to obtain some stability for the results. The data should then be analyzed with structural equation modeling techniques or comparable procedures to identify empirically the optimal scoring device for the latent knowledge monitoring construct.

Actual Knowledge and Estimates of Knowledge

Research has indicated that vocabulary test scores are one of the most powerful predictors of classroom learning (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). KMA scores combine both students' estimates of what they know and their actual knowledge. Thus the ++ score is a composite of both actual word knowledge, determined by the raw score on the vocabulary test, and the students' correct estimates of that knowledge. Each of the studies described above examined whether the KMA estimates contributed independent information beyond that accounted for by students' actual word knowledge. Operationally, this question was analyzed by comparing the variance accounted for by correct estimates (++ and -- combined) with the variance accounted for using only the number correct on the vocabulary test (++ added to --+). Table 7 summarizes these results for each of the studies.

Table 7 indicates that in Studies V, VII, and VIII (four comparisons), actual knowledge alone, determined by raw score on the vocabulary test, accounted for more variance (ranging from 1–17 percent) than did the estimates. Also, there seemed to be little difference between actual knowledge and estimates in Study III. When college students' estimates of their knowledge in Introductory Psychology courses were related to their AP Psychology scores, the effect size for actual knowledge was 13 percent (Study VII) and 17 percent (Study VIII) greater than for estimates of knowledge. When relationships between indices of Introductory Psychology students' in-class performance and KMA scores were analyzed (Study VIII), the effect size for knowledge alone was 6 percent greater.

It is not unusual for knowledge of vocabulary, even in an unrelated domain, to be an important predictor of students' grades in college exams, such as the multiple-choice test and the AP Examination administered in Studies VII and VIII. Vocabulary scores based on words not directly related to a particular course curriculum have been shown to be powerful predictors of all types of classroom learning (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). Thus, findings that such scores were highly related to how much students learned in a psychology course (determined by either the AP Exam or in-class

Table 7

Summary Comparing Knowledge Monitoring Scores (KMA) and Raw Scores

Study

I	KMA accounted for 4% more variance than raw scores.
II	KMA accounted for 5% more variance than raw scores.
III	Correlations similar for KMA and raw scores.
IV	Combined GPA- differentiated KMA scores, effect size = .07, raw scores NS (effect size = .03). English GPA differentiated KMA scores, effect size = .07, raw scores .06. Humanities GPA differentiated KMA scores, effect size = .09, raw scores .05.
V	Vocational high school low and high GPA groups differed on KMA, effect size = .14, and on raw scores, effect size = .16. Predicted score before and after taking test and actual final exam score = <i>ns</i> for KMA and raw scores.
VI	Difference between high school students and dropouts greater on KMA, effect size = .13, than on raw scores, effect size = .10.
VII	AP data and final grade related to KMA, effect size = .14 and raw scores, effect size = .27.
VIII	AP data and final grade related to KMA, effect size = .16, and raw scores, effect size = .33. Class test data related to KMA, effect size = .17, and raw scores, effect size = .23.
IX	KMA r^2 with Metropolitan score = .58, raw score = .27.
X*	
XI	Estimates r^2 with need for feedback = .62, raw score = .04 (<i>ns</i>).
XII	Differences between regular, LD and ADHD Students greater with KMA than raw scores, effect size = .18 compared to .15 for raw scores.

NS = Nonsignificant.

* = Could not be determined.

tests) were not surprising. Furthermore, because students had little prior experience with the content of the AP Examination, they had no basis for estimating their performance on that test. In such instances, it is therefore not unreasonable that actual knowledge may be more important in determining students' achievement than estimates of that knowledge.

Estimates of knowledge accounted for more variance in seven of the studies (nine comparisons, ranging in effect size or r^2 from 1 percent–58 percent, with a median of 4 percent more variance), compared with actual knowledge. The largest differences occurred in the investigation of need for feedback (Study XI), in which the raw score on the vocabulary test accounted for an insignificant 4 percent of the variance, while accurate knowledge monitoring estimates accounted for a highly significant 62 percent of the variance! Of course, that finding should be replicated with larger samples. Nevertheless, it seems reasonable that students' need for feedback should relate more strongly to their estimates of their knowledge rather than to actual knowledge.

Another large difference between the contributions of estimated and actual knowledge scores occurred in Study

IX, one of the math studies. Students' estimates of the number of problems they could solve accounted for 31 percent more variance than did the number of problems actually solved. The findings of Study IX were replicated in Study X; unfortunately, a computer malfunction made it impossible to compare the estimated and actual scores in that investigation. While the math studies clearly need replication, the findings suggest that estimates of knowledge may be more powerful predictors of achievement in that domain than in word knowledge.

One possible reason for the substantial effects in mathematics compared to word knowledge may deal with domain similarity. That is, estimates of knowledge in math were made with respect to content that was highly similar to the types of problems encountered in actual math courses. As indicated previously, the vocabulary words used in many of the studies were not similar to the words presented in actual courses, perhaps leading to somewhat weaker effects. That interpretation is supported by findings in several of the investigations. In Study I, relating actual declarative word knowledge and estimates of that knowledge to reading comprehension, the strongest relationships were found for KMA scores after students had read the passage in which the vocabulary words were defined. That sequence was obviously very similar to the task students face in reading comprehension tests. In addition, it will be recalled that in Study III, social science and science had the lowest relationships with KMA scores, and that in Study III, the effects for social science and science were insignificant. Since the KMA materials were developed to be quite general, they were probably dissimilar to the types of materials with which students are presented in these more technical subjects. These results suggest that the KMA has stronger effects within a domain, rather than across domains. Schraw, Dunkle, Bendixen, and Roedel (1995) found that knowledge monitoring had both domain-specific and domain-general attributes. Further research is needed to clarify the domain-specific and/or domain-general characteristics of the KMA procedure.

Another possible explanation for the more positive results in the studies involving mathematics relates to the perceived difficulty of the subject. Everson, Tobias, Hartman, and Gourgey (1993) found that students perceive mathematics to be the second most difficult subject, right after science. Conceivably, as suggested below, students' estimates of their knowledge in more difficult subjects are less automatic and involve more reflection about their prior experiences than in less difficult subjects. Students' confidence in and/or their anxiety about these subject areas may also affect their estimates. Further research is needed using materials drawn from mathematics, science, and other technical fields to study both this question and the issue of domain specificity.

The KMA Procedure and Degree of Difficulty

Little information about the difficulty of the various vocabulary and math materials was available prior to their use in any of the studies. This may well have contributed to some of the varying results. It seems reasonable that estimates of knowledge based on students' thoughtful consideration of what they know and don't know would be more substantially related to other variables than estimates made more or less automatically. Rapid answers made with little reflection are most likely when students respond to materials that are very easy for them. Wrong estimates based on such relatively automatic responses probably indicate careless errors, rather than a failure to seriously consider the estimate. More difficult materials may also evoke nonreflective responses, since students may feel that they neither know nor care about what the correct answers to such questions are. Items of moderate difficulty, about which students may have partial knowledge that can be extended by exerting some effort, would appear to be most likely to elicit well-considered responses that correctly reflect students' knowledge monitoring ability.

Item difficulty is also of importance in considering the different KMA scores. In the studies described above, of the four scores generated by the KMA procedure, the greatest number of responses fell into the ++ category. It may be assumed that more difficult items would yield more -- and +- responses, increasing the reliability of these items and the likelihood that they could explain more of the variance between students with high knowledge monitoring ability and their less able peers. Furthermore, having more items in the -- category reduces the agreement between estimates and number correct for two reasons: First, such responses represent accurate estimates but no knowledge about the item, and second, more -- items allows for a smaller percentage of ++ items.

In future research, these expectations about the effects of varying degrees of item difficulty should be tested by using items with a previously determined range of difficulty. It could be hypothesized that the most useful metacognitive knowledge estimates are likely to be generated from materials of moderate difficulty, and that more difficult items will increase the difference between the accuracy of knowledge monitoring estimates and the demonstrated knowledge in a domain.

Relationship of the KMA to Metamemory Research

The KMA procedure described in this report is similar to metamemory research on the feeling of knowing (FOK) and judgment of learning (JOL). FOK judgments "occur

during or after acquisition and are judgments about whether a given currently nonrecallable item is known and/or will be remembered on a subsequent retention test....Judgments of learning (JOL) occur during or after acquisition and are predictors about future test performance on currently recallable items” (Nelson & Nahrens, 1990, p. 130). In terms of this definition, for the studies reviewed in this report, students’ estimates of their ability with respect to both the word list and the math problems were similar to JOLs.

FOK research was originated by Hart (1965), who asked general information questions of students who, after failing to recall an item, had to make a judgment regarding their FOK about that item. Finally, they were asked to select an answer from a set of distractors. The procedure has been extended to asking students to guess if they could recall words learned in a paired-associate task (Hart, 1967; Ryan, Petty, & Wentzlaff, 1982). Nelson, Gerler, and Nahrens (1984) also extended the FOK research to students’ ability to relearn, and to tasks in which students were asked to identify perceptual stimuli. Reder and Ritter (1992) investigated whether students opted either to retrieve or re-calculate mathematical problems, and both the time taken to solve the problems and accuracy of the processes were studied. A review of FOK research indicated that “a large number of studies confirmed that (students)...unable to retrieve a solicited item from memory can estimate with above-chance success whether they will be able to recall it in the future, produce it in response to clues, or identify it among distractors.... The standard finding is that the predictive validity of FOK judgments is above-chance, though far from perfect” (Koriat, 1993, p. 609–610).

The FOK and JOL paradigms differ from the present research in a number of ways. First, FOK judgments are typically elicited following a failure to recall, rather than after every administration of material. Second, in FOK or JOL research usually no attempts are made to enable students to improve their knowledge, as was done in some of the studies reviewed in this report. Third, the purposes of metamemory research are to clarify the mechanisms accounting for FOK and JOL, rather than to use the results as measures of metacognitive knowledge monitoring ability to be related to variables of importance in classroom learning.

Suggestions for Further Research

A number of recommendations for further research have already been made; additional suggestions that do not pertain directly to the previous discussion are presented here. The positive findings relating knowledge monitoring ability to need for feedback suggest that studies of similar variables relating the KMA procedure to

processes relevant to classroom learning may be fruitful. For example, forgetting what has been learned in school may be related to knowledge monitoring ability. It could be inferred that students with high knowledge monitoring ability, by having a clear sense of what they know and do not know, may be able to retrieve more prior learning than those who have a less secure grasp of what they know and do not know and who, hence, may have greater difficulty retrieving prior learning. A pilot study of the knowledge monitoring-forgetting relationship provided substantial support for that reasoning and will soon be followed up.

The relationship between knowledge monitoring ability and the effect of distractibility is another fruitful area for investigation. Even though a great deal of anecdotal evidence suggests that students are readily distracted from their studies, it has been surprisingly difficult to divert students in investigations specifically designed for that purpose (Slater, 1968; Tobias, 1973). While part of the problem may be attributable to motivational phenomena, i.e., the interest level of both the primary and distracting materials seems to be important in determining whether students are successfully diverted from their studies (Tobias, 1973), students’ knowledge monitoring ability may also help to determine whether students are distracted. Students with an accurate grasp of their knowledge would be expected to find distractions less disruptive than those with a hazier notion of what they know and do not know.

Research should also be conducted relating knowledge monitoring ability to depth of knowledge processing (Craik & Lockhart, 1972). Students should be able to distinguish between the known and unknown more accurately if the knowledge was processed at a deep rather than shallow level. Deeper processing should enhance students’ knowledge monitoring ability, and it could be predicted that students will make more accurate distinctions between the known and unknown on material that they are induced to process deeply, either by experimental manipulations or instructions, rather than at a shallow level.

Learning in complex domains such as science and engineering, or making diagnoses in medicine or other fields, often requires that students bring substantial amounts of prior learning to bear in order to understand and acquire new knowledge and/or solve problems. Some prior learning may be recalled imperfectly, or may never have been completely mastered during initial acquisition. Students who can accurately distinguish between what they know and don’t know should be at an advantage while working in such domains, since they are more likely to review and try to relearn imperfectly mastered materials needed for particular tasks compared with those who are less accurate in estimating their own knowledge.

Further research is also needed to determine the relationships between the KMA procedure and self-report measures of metacognition, study skills, and self-regulation. These constructs have some similarity to the KMA procedure and positive relationships should be obtained. Finally, the relationship between knowledge monitoring ability and measures of intelligence should be investigated. Sternberg (1991) has suggested that metacognition should be a component of intelligence tests; presumably those who consider metacognition an executive process (Borkowski, Chan and Muthukrishna, in press) would also agree with that recommendation. Research findings (Schraw, in press) indicate that academically able students have higher knowledge monitoring ability than those less able. Therefore, positive relationships between the KMA procedure and measures of general intellectual ability may be expected.

Endnotes

1. Study I was presented at the annual convention of the American Psychological Association, in San Francisco, August 1991. That paper was coauthored by S. Tobias, H. Hartman, H. Everson, H. & A. Gourgey, see References.
2. This study, by Howard Everson, Ivan Smodlaka, and Sigmund Tobias, was published was published in *Stress, Anxiety, and Coping*, 1994, see References.
3. A paper based on Studies III and IV was presented at the annual meeting of the American Educational Research Association in San Francisco, April 1995.
4. The data for this study were collected by Deno Charalambous.
5. The data for this study were collected by Heather Gerrity.
6. The data for this study were collected by Dhalma Rosado. This investigation was presented as part of a paper at the annual convention of the American Educational Research Association in San Francisco, April, 1995.
7. The data for this study were collected by Audrey D'Agostino. The study was part of a paper presented at the annual convention of the American Educational Research Association in New Orleans, April, 1994.
8. This study, conducted by Sigmund Tobias, was published in the *Journal of Educational Psychology*, 1995, see References.
9. The data for this study were collected by Nadia Seignon.
10. The data for this study were collected by Julie Wilson.

References

- Anand, P. G., & S. M. Ross (1987). "Using computer-assisted instruction to personalize arithmetic materials for elementary school children." *Journal of Educational Psychology*, 79, 72-78.
- Ashford, S. J., & L. I. Cummings (1983). "Feedback as an individual resource: Personal strategies creating information." *Organization Behavior and Human Performance*, 32, 370-398.
- Baker, L., & L. Cerro (in press). "Assessing metacognition in children and adults." In G. Schraw (Ed.) *Issues in the measurement of metacognition*. Lincoln, NE: Buros/The University of Nebraska Press.
- Borkowski, J. G. (1995, April). "The assessment of executive functioning." Paper presented at the annual convention of the American Educational Research Association, San Francisco.
- Borkowski, J. G., L. K. S. Chan, & N. Muthukrishna (in press). "A process-oriented model of metacognition and executive functioning." In G. Schraw (Ed.) *Issues in the measurement of metacognition*. Lincoln, NE: Buros/The University of Nebraska Press.
- Bracken, B. A. (1982). "Effect of personalized basal stories on the reading comprehension of fourth-grade poor and average readers." *Contemporary Educational Psychology*, 7, 320-324.
- Breland, H., R. J. Jones, & L. Jenkins (1994). *The College Board vocabulary study* College Board Report No. 94-4. New York: College Entrance Examination Board.
- Brown, A. L. (1980). "Metacognitive development and reading." In R. J. Spiro, B. B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-481). Hillsdale, NJ: Erlbaum.
- Brown, A. L., & J. C. Campione (1986). "Psychological theory and the study of learning disabilities." *American Psychologist*, 14, 1059-1068.
- Butler, D. L., & P. H. Winne (1995). "Feedback and self-regulated learning: A theoretical synthesis." *Review of Educational Research*, 65, 245-281.
- Campione, J. C., A. L. Brown, & M. L. Connell (1989). "Metacognition: On the importance of understanding what you are doing." In R. Charles & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 93-114). Reston, VA: National Council of Teachers of Mathematics.
- Cardell-Elawar, M. (1992). "Effects of teaching metacognitive skills to students with low mathematics ability." *Teaching and Teacher Education*, 8, 109-121.
- Carlson, J. S., & K. H. Wiedl (1992). "Principles of dynamic assessment: The application of a specific model." *Learning and Individual Differences*, 4, 153-166.
- College Board. (1988). *The College Board Technical Manual for the Advanced Placement Program*. New York: College Entrance Examination Board.
- Craik, F. I. M., & R. S. Lockhart (1972). "Levels of processing: A framework for memory research." *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Cronbach, L. J. (1951). "Coefficient alpha and the internal structure of tests." *Psychometrika*, 16, 297-334.
- Delaney, H. D., & S. E. Maxwell (1981). "On using analysis of covariance in repeated measures designs." *Multivariate Behavioral Research*, 16, 105-124.
- Descriptive Test of Language Skills*. (1979). Princeton, NJ: College Entrance Examination Board.
- Douglas, V. T., R. G. Barr, M. E. O'Neil, & B. G. Britton (1986). "Short-term effects of methyphenidate on the cognitive learning and academic performance of children with attention deficit disorder in the laboratory and classroom." *Journal of Child Psychiatry and Psychology*, 27, 191-211.
- Ekstrom, R. B., M. E. Goertz, J. M. Pollack, & D. A. Rock (1986). "Who drops out of high schools and why? Findings from a national study." *Teachers College Record*, 87, 356-372.
- Everson, H. T., H. Hartman, S. Tobias, & A. Gourgey (1991, June). "A metacognitive reading strategies scale: Preliminary validation evidence." Report presented at the annual convention of the American Psychological Association, Washington, D. C.
- Everson, H. T., I. Smolaka, & S. Tobias (1994). "Exploring the relationship of test anxiety and metacognition to reading test performance. A cognitive analysis." *Anxiety, Stress, and Coping*, 7, 85-96.
- Everson, H. T., S. Tobias, H. Hartman, and A. Gourgey (1993). "Anxiety and the curriculum: The subject matters." *Anxiety, Stress, and Coping*, 6, 1-8.
- Eysenck, M. (1988). "Anxiety and attention." *Anxiety Research. An International Journal*, 1, 9-15.
- Fenema, E., & J. A. Sherman (1976). "Fenema-Sherman mathematics attitude scales." *Journal for Research in Mathematics Education*, 7, 324-326.
- Flavell, J. (1979). "Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry." *American Psychologist*, 34, 906-911.
- Green, D. M., & J. A. Swets (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Guthke, J. (1992). "Learning tests: The concept, main research findings, problems and trends." *Learning and Individual Differences*, 4, 137-151.
- Harris, L. J. (1990). "Performance estimation and achievement in introductory psychology." *Perceptual and Motor Skills*, 71, 1072-1074.
- Hart, J. T. (1965). "Memory and the feeling-of-knowing experience." *Journal of Educational Psychology*, 56, 208-216.
- Hart, J. T., (1967). "Memory and the memory monitoring process." *Journal of Verbal Learning and Verbal Behavior*, 6, 685-691.

- Hembree, R. (1988). "Correlates, causes, effects, and treatment of test anxiety." *Review of Educational Research*, 58, 47-78.
- Herndon, J. N. (1987). "Learner interests, achievement, and continuing motivation in instruction." *Journal of Instructional Development*, 10, 11-14.
- Hidi, S. (1990). "Interest and its contribution as a mental resource for learning." *Review of Educational Research*, 60, 549-571.
- Hidi, S., & V. Anderson (1992). "Interest and its differentiated effects on reading and writing." In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 215-238). Hillsdale, NJ: Erlbaum.
- Holen, M. C., & R. C. Newhouse (1976). "Student self-prediction of academic achievement." *Journal of Educational Research*, 69, 219-220.
- Hotelling, H. (1940). "The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters." *Annals of Mathematical Statistics*, 11, 271-283.
- Jacobs, J. F., & S. G. Paris (1987). "Children's metacognition about reading: Issues in definition, measurement, and instruction." *Educational Psychologist*, 22, 255-278.
- Just, M. A., & P. A. Carpenter (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.
- Keefer, K. E. (1971). "Characteristics of students who make accurate and inaccurate self-predictions of college achievement." *Journal of Educational Research*, 64, 401-404.
- Koriat, A. (1993). "How do we know that we know? The accessibility model of the feeling of knowing." *Psychological Review*, 100, 609-639.
- Lester, F. K., J. Garofalo, & D. L. Kroll (1989). "The role of metacognition in mathematical problem solving." Final report to the National Science Foundation. Grant number MDR 85-50346. School of Education, Indiana University, Bloomington, IN.
- Lidz, C. (1992). "Dynamic assessment: Some thoughts on the model, the medium, and the message." *Learning and Individual Differences*, 4, 125-136.
- Lopez, C. L. (1989). "Effects of personalized math instruction for Hispanic students." Paper presented at the annual convention of the American Educational Research Association, San Francisco.
- Lopez, C. L. (1990). "Personalizing math word problems." Proceedings of selected presentations at the convention of the Association for Educational Communications and Technology. (ED 323939).
- Macmillan, N. A., & C. D. Creelman (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Mann, D. (1986). "Can we help dropouts? Thinking about the undoable." *Teachers College Record*, 87, 307-322.
- McKeachie, W. J. (1974). "The decline and fall of the laws of learning." *Educational Researcher*, 3(3), 7-11.
- Meichenbaum, D., S. Burland, L. Gruson, & R. Cameron (1985). "Metacognitive assessment." In S. R. Yussen (Ed.), *The growth of reflection in children* (pp. 3-27). New York: Academic Press.
- Metropolitan Achievement Test. (1985). San Antonio, TX: The Psychological Corporation.
- Montague, M. (1992). "The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities." *Journal of Learning Disabilities*, 25, 230-248.
- Morris, L. W., M. A. Davis, & C. H. Hutchings (1981). "Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale." *Journal of Educational Psychology*, 73, 541-555.
- National Assessment of Educational Problems (1979). *Second assessment of mathematics: Mathematical applications*. Denver, CO: Educational Commission of the States.
- Nelson, T. O., D. Gerler, & L. Nahrens (1984). "Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning." *Journal of Experimental Psychology: General*, 113, 282-300.
- Nelson, T. O., & L. Nahrens (1990). "Metamemory: A theoretical framework and new findings." In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125-173). New York: Academic Press.
- O'Neil, H. F., Jr. (1991). "Metacognition: Teaching and measurement." Paper delivered at the annual convention of the American Psychological Association, August, San Francisco.
- Paris, S. (1988, April). "Fusing skill and will: The integration of cognitive and motivational psychology." Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Pintrich, P. R., D. A. Smith, T. Garcia, & W. J. McKeachie (1991). *A manual for the use of the Motivated Strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., C. A. Wolters, & G. P. Baxter (in press). "Assessing metacognition and self-regulated learning." In G. Schraw (Ed.) *Issues in the measurement of metacognition*. Lincoln, NE: Buros/The University of Nebraska Press.
- Pressley, M. (in press). "Development of grounded theories of complex cognitive processing: Exhaustive within- and between-study analyses of think-aloud data." In G. Schraw (Ed.) *Issues in the measurement of metacognition*. Lincoln, NE: Buros/The University of Nebraska Press.
- Pressley, M., & P. Afflerbach (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Pressley, M., & E. S. Ghatala (1990). "Self-regulated learning: Monitoring learning from text." *Educational Psychologist*, 25, 19-34.

- Reder, L. M., & F. E. Ritter (1992). "What determines initial feeling of knowing? Familiarity with question terms, not with answers." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 435-451.
- Renninger, K. A., S. Hidi, & A. Krapp (1992). *The role of interest in learning and development*. Hillsdale, NJ: Erlbaum.
- Ross, S. M., & P. G. Anand (1987). "A computer-based strategy for personalizing verbal problems in teaching mathematics." *Educational Communication and Technology Journal*, 35, 151-162.
- Royer, J. M., M. S. Carlo, R. Dufresne, & R. Mestre (1994). "The assessment of levels of expertise while reading." Amherst, MA: University of Massachusetts. Unpublished document.
- Royer, J. M., C. A. Cisero, & M. Carlo (1993). "Techniques and procedures for assessing cognitive skills." *Review of Educational Research*, 63, 201-243.
- Royer, J. M., D. J. Lynch, R. K. Hambleton, & C. Bulgarelli (1984). "Using the sentence verification technique to assess the comprehension of technical text as a function of level of expertise." *American Educational Research Journal*, 21, 839-869.
- Ryan, M. P., C. R. Petty, & J. R. Wentzlaff (1982). "Motivated remembering efforts during tip-of-the-tongue states." *Acta Psychologica*, 51, 137-147.
- Sarason, I. G. (1987). "Test anxiety, cognitive interference, and performance." In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction* (Vol. 3): *Conative and affective process analyses* (pp. 131-142). Hillsdale, NJ: Erlbaum.
- Schiefele, U. (1990). "The influence of topic interest, prior knowledge, and cognitive capabilities on text comprehension." In J. M. Pieters, K. Breuer, & P. R. J. Simons (Eds.) *Learning environments* (pp. 323-337). Berlin: Springer.
- Schiefele, U. (1991). "Interest, learning, and motivation." *Educational Psychologist*, 26, 299-323.
- Schiefele, U. (1992a). "Topic interest and levels of text comprehension." In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- Schiefele, U. (1992b). "Content related motivation and learning from text." Paper presented at the annual convention of the American Educational Research Association, San Francisco, CA, April.
- Schoenfeld, A. (1992). "Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics." In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York: Macmillan.
- Schraw, G. (in press). *Issues in the measurement of metacognition*. Lincoln, NE: Buros/The University of Nebraska Press.
- Schraw, G., & R. S. Dennison (1994). "Assessing metacognitive awareness." *Contemporary Educational Psychology*, 19, 460-475.
- Schraw, G., M. E. Dunkle, L. D. Bendixen, & T. D. Roedel (1995). "Does a general monitoring skill exist?" *Journal of Educational Psychology*, 87, 433-444.
- Slater, B. R. (1968). "Effects of noise on pupil performance." *Journal of Educational Psychology*, 59, 239-243.
- Spielberger, C. D., H. P. Gonzalez, C. J. Taylor, W. D. Anton, B. Algaze, G. R. Ross, & L. G. Westberry (1980). *Preliminary professional manual for the test anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- SPSS. (1993). *SPSS for windows: Release 6*. Chicago, IL: SPSS.
- Sternberg, R. J. (1991). "Toward better intelligence tests." In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 31-39). Englewood Cliffs, NJ: Prentice-Hall.
- Swanson, H. L., & M. F. Trahan (1992). "Learning disabled readers' comprehension of computer mediated text: The influence of working memory, metacognition, and attribution." *Learning Disabilities Research and Practice*, 7, 74-86.
- Tanner, J. T. (1990). "Reluctant rebels. A case study of Edmonton high school dropouts." *Canadian Review of Sociology and Anthropology*, 27, 74-93.
- Tobias, S. (1969). "Effect of creativity, response mode, and subject matter familiarity on achievement from programmed instruction." *Journal of Educational Psychology*, 60, 453-460.
- Tobias, S. (1973). "Distraction, response mode, anxiety, and achievement from CAI." *Journal of Educational Psychology*, 65, 233-237.
- Tobias, S. (1985). "Test anxiety: Interference, defective skills, and cognitive capacity." *Educational Psychologist*, 20, 135-142.
- Tobias, S. (1989). "Another look at research on the adaptation of instruction to student characteristics." *Educational Psychologist*, 24, 213-227.
- Tobias, S. (1992). "The impact of test anxiety on cognition in school learning." In K. Hagtvet (Ed.), *Advances in Test Anxiety Research* (Vol. 7, pp. 18-31). Lisse, Netherlands: Swets & Zeitlinger.
- Tobias, S. (1994a). "Interest, prior knowledge, and learning." *Review of Educational Research*, 64, 37-54.
- Tobias, S. (1994b, April). "Interest and metacognition in word knowledge and mathematics." Paper presented at the annual convention of the American Educational Research Association, New Orleans, LA.
- Tobias, S. (1995). "Interest and metacognitive word knowledge." *Journal of Educational Psychology*, 87, 399-405.
- Tobias, S., H. Hartman, H. T. Everson, & A. Gourgey (1991, August). "The development of a group-administered, objectively scored metacognitive evaluation procedure." Paper presented at the annual convention of the American Psychological Association, San Francisco.
- Touchstone Applied Sciences Associates. (1991). *Degrees of Reading Power*, New York: Touchstone.

-
- Tuckman, B., & T. L. Sexton (1992). "The effects of informational feedback and self-beliefs on the motivation to perform a self-regulated task." *Journal of Research in Personality*, 26, 121-127.
- Van Haneghan, J. P., & L. Baker (1989). "Cognitive monitoring in mathematics." In C. B. McCormick, G. Miller, & M. Pressley (Eds.), *Cognitive strategy research: From basic research to educational applications* (pp. 215-238). New York: Springer Verlag.
- Werts, C., R. L. Linn, & K. G. Joreskog (1978). "Reliability of college grades from longitudinal data." *Educational and Psychological Measurement*, 38, 89-95.
- Westby, C. E., & S. K. Cutler (1994). "Language and ADHD: Understanding the bases and treatment of self-regulatory deficits." *Topics in Language Disorders*, 14, 58-76.
- Willingham, W. W., C. Lewis, R. Morgan, & L. Ramist (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.
- Wright, J. P., & C. D. Wright (1986). "Personalized verbal problems: An application of the language experience approach." *Journal of Educational Research*, 79, 358-362.